

istat working papers

N.1
2021

Il supporto dell'Istat nella progettazione del 4° Censimento della Popolazione in Etiopia: sfide e opportunità dell'Indagine di Controllo della Copertura

Giancarlo Carbonetti, Paolo Giacomi, Filomena Grassia e Alessandra Nuccitelli

Direttrice Responsabile:

Patrizia Cacioli

Comitato Scientifico**Presidente:**

Gian Carlo Blangiardo

Componenti:

Corrado Bonifazi	Vittoria Buratta	Ray Chambers	Francesco Maria Chelli
Daniela Cocchi	Giovanni Corrao	Sandro Cruciani	Luca De Benedictis
Gustavo De Santis	Luigi Fabbris	Piero Demetrio Falorsi	Patrizia Farina
Jean-Paul Fitoussi	Maurizio Franzini	Saverio Gazzelloni	Giorgia Giovannetti
Maurizio Lenzerini	Vincenzo Lo Moro	Stefano Menghinello	Roberto Monducci
Gian Paolo Oneto	Roberta Pace	Alessandra Petrucci	Monica Pratesi
Michele Raitano	Maria Giovanna Ranalli	Aldo Rosano	Laura Terzera
Li-Chun Zhang			

Comitato di redazione**Coordinatrice:**

Nadia Mignolli

Componenti:

Ciro Baldi	Patrizia Balzano	Federico Benassi	Giancarlo Bruno
Tania Cappadozzi	Anna Maria Cecchini	Annalisa Cicerchia	Patrizia Collesi
Roberto Colotti	Stefano Costa	Valeria De Martino	Roberta De Santis
Alessandro Faramondi	Francesca Ferrante	Maria Teresa Fiocca	Romina Fraboni
Luisa Franconi	Antonella Guarneri	Anita Guelfi	Fabio Lipizzi
Filippo Moauro	Filippo Oropallo	Alessandro Pallara	Laura Peci
Federica Pintaldi	Maria Rosaria Prisco	Francesca Scambia	Mauro Scanu
Isabella Siciliani	Marina Signore	Francesca Tiero	Angelica Tudini
Francesca Vannucchi	Claudio Vicarelli	Anna Villa	

Supporto alla cura editoriale:

Vittorio Cioncoloni

Istat Working Papers

Il supporto dell'Istat nella progettazione del 4° Censimento della Popolazione in Etiopia: sfide e opportunità dell'Indagine di Controllo della Copertura

N. 1/2021

ISBN 978-88-458-2043-4

© 2021

Istituto Nazionale di Statistica - Istat

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.

Il supporto dell'Istat nella progettazione del 4° Censimento della Popolazione in Etiopia: sfide e opportunità dell'Indagine di Controllo della Copertura

Giancarlo Carbonetti, Paolo Giacomi, Filomena Grassia e Alessandra Nuccitelli¹

Sommario

L'Istat ha fornito supporto metodologico e assistenza tecnica alla Central Statistical Agency dell'Etiopia per la preparazione del 4° Censimento della Popolazione nell'ambito di un progetto di cooperazione finanziato dal governo italiano tramite l'Agenzia Italiana per la Cooperazione allo Sviluppo, per la durata di 24 mesi.

Il lavoro riassume le principali attività svolte nell'ambito della componente del progetto dedicata alla progettazione dell'Indagine di Controllo della Copertura. In particolare, si discutono alcuni aspetti relativi alla strategia e alla fattibilità dell'indagine in un contesto caratterizzato da elementi di forti criticità, quali la carenza di infrastrutture, la mancanza di una base anagrafica di partenza e la presenza di forti flussi migratori interni.

Inoltre, il lavoro riporta i risultati principali delle prime due indagini pilota e alcune linee guida per l'utilizzo di un'applicazione Web, appositamente sviluppata per effettuare l'abbinamento dei dati di indagine con i dati censuari.

Parole chiave: abbinamento dei dati, campionamento, cooperazione internazionale, Post-Enumeration Survey, raccolta dei dati.

Abstract

Istat provided methodological support and technical assistance to the Central Statistical Agency of Ethiopia for the preparation of the 4th Population Census, as part of a 24-month cooperation project financed by the Italian government through the Italian Agency for Development Cooperation.

The work summarises the main activities carried out within the project component aimed at designing the Post-Enumeration Survey. In particular, some aspects related to the strategy and feasibility of the survey in a context characterised by highly critical elements – such as the lack of infrastructure, the absence of a population register and the presence of strong internal migration flows – are discussed.

Moreover, the work reports the main findings from the first two pilot surveys and some guidelines for the use of a Web-based application, specifically developed to link Census and PES data.

Keywords: data matching, sampling, international cooperation, Post-Enumeration Survey, data collection.

¹ carbonet@istat.it; giacomi@istat.it; grassia@istat.it; nuccitel@istat.it.

I punti di vista espressi sono quelli degli autori e non riflettono necessariamente le opinioni ufficiali dell'Istituto Nazionale di Statistica - Istat.

Indice

	Pag.
1. Introduzione	5
2. Quadro di riferimento	6
3. Il contesto nazionale	9
4. Obiettivi della PES	10
5. Disegno di campionamento	11
6. Raccolta dei dati	12
6.1 Tecniche di rilevazione	12
6.2 Questionario	12
6.3 Rilevazione sul campo	13
7. Processo di lavorazione dei dati	14
7.1 Trattamento preliminare dei dati	14
7.2 Abbinamento dei dati	15
8. Procedimento di stima	17
9. Esperienze e risultati acquisiti dalle prime due indagini pilota	19
10. Considerazioni conclusive	21
Appendice	23
Riferimenti bibliografici	33

1. Introduzione

A partire dagli anni 90 del secolo scorso l'Istat ha visto crescere progressivamente il suo impegno nell'attività di cooperazione internazionale che si concretizza nella realizzazione di programmi o progetti di *capacity building* rivolti a specifiche regioni o Paesi di interesse principalmente nazionale e comunitario. Tali progetti sono finalizzati allo sviluppo di moderni sistemi statistici attraverso l'assistenza tecnica agli Istituti nazionali di statistica per il rafforzamento del loro capitale umano, scientifico, tecnologico, organizzativo ed istituzionale.

In Etiopia, l'Istat ha contribuito al rafforzamento dell'Agenzia centrale di statistica (Central Statistical Agency – CSA) fornendo supporto tecnico e metodologico per la preparazione del 4° Censimento della Popolazione e delle Abitazioni, nell'ambito di un intervento di 24 mesi finanziato dall'Agenzia Italiana per la Cooperazione allo Sviluppo (AICS)².

La complessità dell'intervento in Etiopia è derivata anche dall'introduzione di una serie di innovazioni metodologiche e tecnologiche che, in linea con gli orientamenti internazionali, hanno riguardato:

- la transizione a mappe completamente digitali, utilizzando la tecnologia mobile GIS (Geographic Information System);
- il passaggio alla rilevazione *paperless*, attraverso l'impiego di dispositivi portatili (*tablet*) con sistema operativo Android;
- la progettazione della *Post-Enumeration Survey* (PES).

Tali innovazioni sono state messe in esercizio nelle varie indagini pilota condotte durante il progetto di cooperazione: tre per il Censimento e tre per la PES.

Il lavoro riassume le principali attività svolte congiuntamente dall'Istat e dalla CSA nell'ambito della componente del progetto “Support PES design” e le indicazioni metodologiche più rilevanti fornite per la preparazione e l'implementazione della PES in un contesto caratterizzato da elementi di forti criticità, quali la carenza di infrastrutture, l'assenza di una base anagrafica di partenza e la presenza di forti flussi migratori interni.

In Etiopia, per i passati Censimenti, questa indagine non è stata mai realizzata. Trattandosi della prima esperienza in questo ambito, la CSA ha ritenuto essenziale uno specifico supporto tecnico da parte della comunità internazionale. In particolare, sulla preparazione della PES è intervenuto anche un gruppo di esperti dell'ONS (Office for National Statistics) con cui il team Istat ha concordato la programmazione delle missioni e i contenuti delle principali raccomandazioni tecniche.

Un quadro riassuntivo del supporto fornito dall'Istat e delle questioni chiave affrontate con la CSA e con i diversi *stakeholder* nazionali e internazionali nel corso del progetto è presentato nel paragrafo 2.

Dopo una breve introduzione alla geografia amministrativa del Paese (paragrafo 3) e agli obiettivi della PES nel contesto specifico (paragrafo 4), vengono discussi alcuni aspetti relativi alla strategia e alla fattibilità dell'indagine in Etiopia, soprattutto con riferimento ai seguenti argomenti³:

- disegno di campionamento (paragrafo 5);
- raccolta dei dati (paragrafo 6);
- trattamento dei dati (paragrafo 7);
- procedimento di stima (paragrafo 8).

La maggior parte delle considerazioni si basa sull'esperienza acquisita in occasione delle prime

² Progetto: “Assistenza nel settore statistico – Censimento della Popolazione” (IATI ID: # XM-DAC-6-4-010649-01-5); periodo di realizzazione dell'intervento: giugno 2016 - giugno 2018.

Gli autori del presente lavoro ringraziano tutti i colleghi della *Central Statistical Agency* coinvolti nella preparazione del Censimento per la fiducia dimostrata e la disponibilità al lavoro di squadra che hanno reso possibile il raggiungimento degli obiettivi del progetto.

Un ringraziamento speciale, inoltre, va anche a Ginevra Letizia – direttore della sede AICS di Addis Abeba nel periodo di realizzazione del progetto – e a Simona Leali – coordinatore del programma AICS in Etiopia “*Child Protection and Gender*” – che hanno costantemente incoraggiato e sostenuto, con autorevolezza ed entusiasmo, il lavoro del team Istat e con le quali è stata condivisa la ricerca di soluzioni operative adatte al contesto locale.

³ Per tutti gli ulteriori dettagli e considerazioni di carattere metodologico/operativo inerenti alla progettazione e all'implementazione di una PES, non affrontati nel presente lavoro, si rimanda a United Nations Secretariat (2010) e a United Nations (2017).

due indagini pilota della PES condotte durante il periodo di attuazione del progetto (rispettivamente, nei mesi di febbraio e settembre 2017). Vale la pena evidenziare che il team Istat, pur avendo contribuito attivamente alla progettazione della terza indagine pilota della PES, non ha potuto valutarne appieno i risultati, in quanto la rilevazione dei dati si è conclusa immediatamente dopo la scadenza del progetto (giugno 2018).

Le principali evidenze emerse dalle prime due indagini pilota sono riportate nel paragrafo 9; vengono poi tratte alcune considerazioni conclusive in merito alla fattibilità della PES in Etiopia e all'utilizzazione dei risultati finali, nel caso in cui l'indagine venisse effettivamente condotta⁴ (paragrafo 10).

Infine, l'appendice include alcune linee guida per l'uso di un'applicazione Web prototipale, appositamente sviluppata dall'Istat per effettuare l'abbinamento esatto⁵ (o *record linkage*) dei dati elementari della PES con quelli del Censimento.

2. Quadro di riferimento

La visione politica della cooperazione allo sviluppo dell'Italia riconosce il rafforzamento delle capacità statistiche come un'area strategica prioritaria, considerata la necessità di dati affidabili e aggiornati per la formulazione delle politiche di sviluppo e per la *good governance* di ogni Paese, nonché per misurare il raggiungimento degli Obiettivi di Sviluppo Sostenibile dell'Agenda 2030 delle Nazioni Unite. Il sostegno dell'Italia al miglioramento delle statistiche ufficiali nei Paesi partner è finanziato dal Ministero degli Affari Esteri e della Cooperazione Internazionale (MAECI) attraverso l'AICS, di cui l'Istat è la principale controparte istituzionale nella realizzazione dei relativi programmi e progetti di assistenza tecnica.

L'approccio italiano è ispirato ai quattro principi⁶ del Partenariato Globale per l'Efficacia della Cooperazione allo Sviluppo e mira ad assicurare una partecipazione effettiva dei Paesi partner in tutte le fasi del ciclo di vita degli interventi di assistenza tecnica; le istituzioni locali sono chiamate ad analizzare la propria situazione attuale, ad attribuire responsabilità e a lavorare insieme per trovare soluzioni sostenibili nel medio-lungo termine.

In quest'ottica, la coerenza con le strategie nazionali per lo sviluppo delle statistiche e con i piani di sviluppo nazionali dei Paesi è sempre ricercata nella progettazione e nell'attuazione delle attività. Sul piano operativo, l'appropriazione dei risultati, delle innovazioni introdotte e dei miglioramenti ottenuti è realizzata mettendo al centro dell'azione le persone, coinvolgendo i diversi livelli dell'organizzazione e del sistema statistico, potenziando i processi interni senza proporre soluzioni preconfezionate e importate dall'esterno, ma adattandole allo specifico contesto (Amarone *et al.*, 2016).

Questi principi hanno guidato anche la realizzazione del progetto in Etiopia, originato nel 2015 da una richiesta di supporto da parte del governo etiope. Tale richiesta è stata accolta dal governo italiano in linea con le priorità dichiarate per il triennio 2016-2018, che vedevano l'Etiopia come uno dei Paesi partner chiave dell'Africa subsahariana e i censimenti come una delle aree su cui concentrare il sostegno nell'area statistica⁷.

Dal punto di vista del Paese partner, il quadro normativo di riferimento è delineato dalla Carta

⁴ Si precisa che nel momento in cui si scrive il 4° Censimento della Popolazione e delle Abitazioni – originariamente previsto per il 2017 – non è stato ancora realizzato, avendo subito diversi slittamenti anche a causa di motivi legati alla complessa situazione politica che il Paese sta attraversando.

⁵ Il termine “abbinamento esatto” si riferisce all'uso di tecniche algoritmiche per identificare *record* relativi ad una stessa unità statistica e contenuti in archivi diversi.

⁶ I quattro principi sono *ownership* dei Paesi partner, focus su risultati, partenariati inclusivi, trasparenza e responsabilità reciproca sui progressi realizzati (ulteriori dettagli si trovano al seguente indirizzo: <https://www.oecd.org/dac/effectiveness/latestdocuments/>).

⁷ Per le priorità della cooperazione italiana nel triennio 2016-2018 si rinvia ai documenti di programmazione del MAECI disponibili ai seguenti indirizzi:

https://www.aics.gov.it/wp-content/uploads/2017/02/DGCS_La_statistica_come_conoscenza_2015.pdf

https://www.aics.gov.it/wp-content/uploads/2017/04/Programmazione_triennale_2016-2018_approvato.pdf.

Costituzionale della Repubblica Democratica Federale dell’Etiopia promulgata nel 1995 – che stabilisce lo svolgimento di un Censimento della Popolazione una volta ogni dieci anni – e dalla proclamazione n° 442/2005 che, nell’istituire la CSA, ha conferito all’Agenzia stessa il compito di raccogliere, elaborare e diffondere le necessarie informazioni socio-economiche e demografiche attraverso censimenti, indagini campionarie e sistemi di gestione dei dati di fonte amministrativa⁸.

La programmazione strategica per il settore statistico (Second National Strategy for the Development of Statistics – NSDSII) riferita al quinquennio 2015/16-2019/20 aveva, inoltre, incluso il Censimento della Popolazione e delle Abitazioni tra le attività strategiche prioritarie, identificando i punti di forza e debolezza del contesto in cui l’esercizio censuario si sarebbe inserito e le azioni da intraprendere per il miglioramento dell’intero Sistema Statistico Nazionale⁹.

Per quanto riguarda la CSA, al momento di identificazione del progetto risultava essere un centro di eccellenza nazionale con:

- una robusta struttura organizzativa dotata di uffici periferici in tutte le nove regioni dello Stato, a garanzia della capillare raccolta dei dati sul territorio;
- una consolidata reputazione in termini di imparzialità e qualità delle statistiche prodotte, che la collocava tra i più avanzati Istituti nazionali di statistica dell’Africa subsahariana.

Tuttavia, nello svolgimento dei propri compiti istituzionali, l’Agenzia si trovava a fronteggiare una serie di difficoltà legate principalmente:

- all’inadeguatezza della dotazione di personale – concentrato principalmente nella rete degli uffici territoriali – derivante da un elevato turnover, a causa del basso livello dei salari e della mancanza di percorsi di carriera;
- all’insufficienza dei fondi allocati alla funzione statistica, che rendeva difficile lo sviluppo olistico dei processi e la modernizzazione delle infrastrutture a supporto della produzione.

In particolare, è stato riscontrato che, a causa dell’elevato ricambio del personale della CSA, nessuno dello staff incaricato della gestione dei processi censuari era stato coinvolto nell’attuazione del Censimento precedente. Tale circostanza ha reso estremamente urgente il bisogno di formazione specifica, soprattutto nelle aree più direttamente toccate dall’introduzione delle innovazioni di processo.

D’altro canto, l’iniziale mancanza di fondi governativi per la copertura dell’intero budget del Censimento e, in particolare, per l’acquisizione dei *tablet* e delle relative *power bank* per la raccolta dei dati, ha reso ancora più evidente la necessità di un supporto tecnico e finanziario da parte della comunità dei donatori e partner internazionali nell’impegno per lo sviluppo, tra cui World Bank, DFID (UK Department for International Development), UNFPA (United Nations Population Fund) e USCB (United States Census Bureau).

Il supporto italiano all’Etiopia si è inserito nel quadro delle complesse e articolate attività di preparazione del 4° Censimento della Popolazione e delle Abitazioni, cui hanno contribuito, con differenti modalità, i diversi donatori e partner internazionali appena citati.

Il forte coordinamento internazionale, a livello istituzionale e tra i diversi gruppi di esperti, ha garantito la complementarità dei singoli interventi di assistenza tecnica e l’integrazione tra i processi posti in essere con il supporto di ciascun progetto, incrementando l’efficacia delle azioni e consentendo lo sviluppo sistematico delle abilità e conoscenze da parte dello staff della CSA.

In questo contesto, il progetto italiano è stato realizzato nell’arco temporale di due anni, da giugno 2016 a giugno 2018, e ha seguito due principali direttrici, l’una finalizzata alla predisposizione dell’infrastruttura informatica per il Censimento e l’altra orientata a sviluppare competenze nell’area metodologica e dell’analisi dei dati.

In ambito informatico, l’assistenza tecnica ha avuto come obiettivo lo sviluppo di un sistema generalizzato di monitoraggio della rilevazione *metadata driven*, integrato con il sistema di *data collection* CSPro realizzato dall’USCB (Bruno *et al.*, 2020). All’interno di questa componente è stato

⁸ Alla CSA è anche attribuito un ruolo di coordinamento all’interno del Sistema Statistico Nazionale, incluso il compito di fornire guida tecnica e assistenza alle agenzie e istituzioni governative (Ministeri, Dipartimenti, Agenzie – MDA) per la realizzazione di moderni sistemi di registrazione dei dati amministrativi e di supportare le medesime MDA nello sviluppo delle necessarie capacità. Per approfondimenti si rinvia al testo integrale della proclamazione n° 442/2005, disponibile all’indirizzo: <https://unstats.un.org/unsd/vitalstatkb/KnowledgebaseArticle50879.aspx>.

⁹ La NSDSII per il quinquennio 2015/16-2019/20 è all’indirizzo: <http://www.paris21.org/sites/default/files/Ethiopia-NSDS.pdf>.

fornito anche supporto per la configurazione e il test del *Data Center* dedicato al Censimento.

La componente metodologica si è invece concentrata prevalentemente sulla progettazione della PES e sullo sviluppo di strumenti informatici per il *record linkage*, come descritto nel seguito del presente lavoro.

L'assistenza tecnica è stata fornita essenzialmente attraverso missioni presso la CSA di un gruppo di nove esperti dell'Istat, selezionati internamente tenendo conto delle professionalità richieste per realizzare le diverse linee di attività stabilite dal piano operativo generale, concordato in fase di avvio del progetto e approvato dall'ente finanziatore. Per ciascuna componente del progetto è stato identificato un coordinatore tecnico. I coordinatori hanno supportato il *project leader* nelle scelte di natura tecnica, garantendo continuità e coordinamento delle attività all'interno delle singole componenti e tra le componenti.

Particolare attenzione è stata poi prestata per garantire il continuo coinvolgimento della CSA ai diversi livelli istituzionali, considerato che l'impegno della dirigenza e la disponibilità di un adeguato staff sono essenziali per garantire efficacia, proprietà e ripetibilità dei risultati.

A tal fine, è stata posta in essere una struttura organizzativa "a specchio", con l'individuazione di controparti locali per ciascuna delle figure chiave del progetto (*project leader*, coordinatori di componente, responsabili di specifiche linee di attività), favorendo la corretta circolazione delle informazioni e il tempestivo accordo su contenuti e modalità di lavoro.

La dirigenza della CSA è stata sempre chiamata a partecipare alle riunioni plenarie organizzate in apertura e chiusura di ciascuna missione con lo scopo di fornire aggiornamenti sui risultati raggiunti, su eventuali criticità, rischi e relative azioni di contenimento.

Inoltre, la CSA è stata continuamente incoraggiata a identificare una squadra di esperti (quadri intermedi, esperti di tecnologia dell'informazione e GIS, metodologi e statistici) da coinvolgere stabilmente nelle attività censuarie e nel processo di rafforzamento delle necessarie capacità.

Per ogni missione, sono stati concordati in anticipo obiettivi, risultati tangibili, controparti da coinvolgere e attività da svolgere, secondo quanto previsto dal piano operativo generale.

In linea con l'approccio generale alla cooperazione internazionale adottato dall'Istat, l'assistenza tecnica fornita alla CSA è stata costantemente orientata al rafforzamento delle capacità – a livello individuale e istituzionale – che, sul piano teorico, rappresenta il legame tra i risultati del progetto e la sostenibilità, per assicurare che i miglioramenti apportati siano assorbiti e mantenuti oltre la durata del progetto.

Sul piano pratico, questo risultato è stato raggiunto:

- ponendo una significativa enfasi sul lavorare insieme attraverso attività di *coaching* e *advice*, volte a trasferire il know-how degli esperti Istat alle loro controparti locali;
- introducendo nuove metodologie e innovazioni nei processi censuari, in linea con le raccomandazioni internazionali e a partire da una valutazione degli specifici punti di forza e di debolezza dei sistemi già esistenti.

Il processo di sviluppo e rafforzamento della capacità istituzionale si è basato non solo sull'introduzione di metodologie e sullo sviluppo di strumenti informatici appropriati al contesto nazionale e in linea con le soluzioni proposte dagli altri partner internazionali, ma anche su una forte componente di formazione ad hoc, che ha combinato training teorici su standard e concetti di base con training specifici sull'utilizzo e la manutenzione degli strumenti sviluppati per il Censimento.

Sempre al fine di garantire la sostenibilità dei risultati, alla fine di ogni missione gli esperti dell'Istat hanno rilasciato un rapporto tecnico, comprensivo di raccomandazioni scritte, materiale utilizzato durante la missione e altra eventuale documentazione di supporto. Ciascun rapporto tecnico ha anche incluso la cosiddetta *to-do-list*, ossia la lista dei compiti da svolgere da entrambe le parti, con l'indicazione delle persone responsabili del completamento dell'attività, dei tempi di consegna e delle modalità di scambio con la controparte.

La *to-do-list* ha garantito la continuità delle attività nel tempo e ha guidato il supporto a distanza da parte degli esperti dell'Istat alle controparti etiopi tra una missione e l'altra. Il supporto a distanza si è svolto non solo via e-mail, ma anche utilizzando piattaforme di collaborazione e strumenti specifici per testare/aggiornare gli strumenti informatici sviluppati.

Inoltre, per tutte le innovazioni metodologiche e tecnologiche introdotte durante il progetto, sono stati rilasciati manuali utente e linee guida allo scopo di agevolarne il riuso in altri processi statistici della CSA e di fornire riferimenti anche allo staff non direttamente coinvolto nel progetto.

Oltre a ciò, è stata costantemente ricercata la cooperazione con gli altri partner internazionali impegnati a supportare la preparazione del Censimento, sia per scambiare esperienze e creare sinergie, sia per evitare di sovraccaricare il personale della CSA con attività concomitanti e potenzialmente contrastanti.

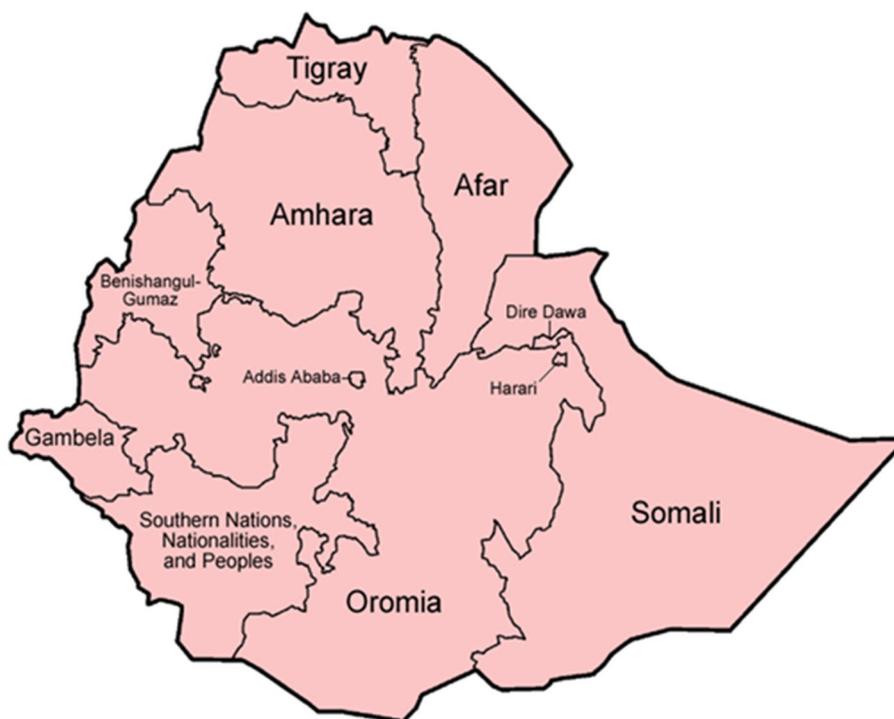
3. Il contesto nazionale

Dal 1995 l’Etiopia è una repubblica parlamentare federale suddivisa in nove regioni federate, etnicamente e politicamente autonome, cui si aggiungono le due città amministrative di Addis Abeba e Harari (Figura 1). Le regioni sono ripartite in zone, a loro volta costituite da distretti i cui confini sono poco stabili nel corso del tempo.

I distretti corrispondono in linea di massima ad unità territoriali chiamate *wereda*, i quali sono composti da *kebele*, associazioni di quartiere che rappresentano le unità amministrative più piccole del governo locale. Ogni *kebele* include circa 500 famiglie o l’equivalente di 3.500-4.000 individui. Ciascun centro abitato di almeno 2.000 persone contiene almeno un *kebele*; all’interno dei *kebele* sono disegnate le sezioni di Censimento (mediamente tre o quattro).

I limiti delle unità amministrative *wereda* e *kebele* sono decisi dai governi locali e soggetti a continui cambiamenti in relazione a condizioni socio-demografiche, abitative e ambientali.

Figura 1. Mappa delle regioni dell’Etiopia.



Dal punto di vista demografico, l’Etiopia è il secondo Paese più popolato dell’Africa subsahariana. Secondo stime della CSA, nel 2013 in Etiopia risiedevano circa 86 milioni di persone, con un tasso di crescita della popolazione del 2,6 per cento annuo (tra i più elevati al mondo).

La rapida crescita della popolazione mette a dura prova le risorse naturali disponibili. Oltre a importanti flussi migratori interni, guidati da fattori economici, climatici e politici (siccità, guerre,

povertà, disordini e rivolte, processi di reinsediamento o villaggizzazione forzata), l'Etiopia registra il più alto numero di rifugiati dell'intero continente africano. Questo dato va messo in relazione con la posizione strategica del Paese, che è al contempo zona di origine e transito di migranti diretti verso le principali destinazioni europee, ma anche nel Golfo, in Medio Oriente e Sud Africa.

Vale la pena evidenziare che l'Etiopia è caratterizzata da un notevole pluralismo religioso, con l'ortodossia che rimane la fede più praticata, incalzata dall'Islam e, in misura ridotta, da altre minoranze.

Inoltre, la parcellizzazione etnica è molto alta (più di 80 gruppi), così come la prevalenza rurale. Secondo l'ultimo Censimento demografico del 2007, l'Etiopia era uno dei Paesi meno urbanizzati al mondo, con il 16,4 per cento della popolazione che risiedeva in contesti urbani contro l'83,6 per cento che viveva nelle zone rurali.

4. Obiettivi della PES

Il Censimento della Popolazione costituisce la rilevazione più ampia e costosa che un Paese possa intraprendere e i suoi risultati vengono utilizzati sempre più spesso per scopi politici e di pianificazione. È pertanto importante conoscere la qualità dei risultati finali cercando di valutare il tipo e l'entità degli errori introdotti nel corso della rilevazione.

La presenza di inesattezze nei dati non diminuisce l'importanza del Censimento nella misura in cui gli utenti sono consapevoli dei limiti dei risultati e gli errori non influiscono pesantemente sui principali utilizzi delle statistiche finali.

Tra i diversi metodi che sono stati sviluppati per valutare la qualità di un Censimento, rientrano quelli basati sull'analisi demografica, sul confronto dei totali censuari con i dati di altre fonti, sulla verifica della corrispondenza (abbinamento) tra le risposte date al Censimento e le risposte fornite in occasione di un'indagine di controllo post-censuaria (PES).

Soprattutto nei Paesi in cui manca un sistema evoluto di registri della popolazione e le indagini demografiche non sono effettuate con regolarità, lo svolgimento di una PES sembra essere l'unica strada percorribile per valutare i risultati del Censimento, vista la scarsità di dati completi e attendibili provenienti da fonti alternative e necessari per l'applicazione di altri metodi¹⁰.

Fondamentalmente, una PES consiste in una ripetizione delle operazioni di enumerazione su un campione rappresentativo della popolazione di interesse per la rilevazione censuaria.

I risultati dell'abbinamento esatto dei dati elementari rilevati dalla PES con quelli del Censimento permettono di valutare l'incidenza degli errori di copertura e/o di misura generati dall'operazione censuaria.

Mentre gli errori di copertura sono dovuti a omissioni o a inclusioni errate di individui e/o unità abitative durante lo svolgimento del Censimento, gli errori di misura si verificano quando informazioni su caratteristiche di persone e/o unità abitative rilevate attraverso il questionario del Censimento sono riferite o riportate in modo non conforme al vero (ad esempio, le informazioni su *Sesso, Età, Stato civile*). Gli errori di copertura possono condurre a un conteggio della popolazione distorto, per eccesso o per difetto¹¹.

Oltre alla valutazione degli errori di copertura e/o di misura, i risultati di una PES possono consentire di rettificare i risultati del Censimento o anche far emergere metodologie o operazioni che, una volta implementate, hanno dato luogo a risultati non soddisfacenti (ad esempio, problemi nella rete dei rilevatori o nella definizione di alcune sezioni di Censimento possono essere evidenziati dall'omissione di intere unità abitative).

È possibile progettare la PES in modo tale da ottenere stime affidabili degli errori, non solo a livello nazionale, ma anche per specifiche sottoaree geografiche e/o per una serie di caratteristiche

¹⁰ In particolare, vale la pena sottolineare che nei Paesi meno sviluppati la carenza di dati attendibili sulle migrazioni internazionali rende discutibile l'utilizzo delle analisi demografiche per valutare la qualità del Censimento della Popolazione.

¹¹ È importante precisare che gli errori di copertura si definiscono tali in relazione all'unità territoriale utilizzata per riferire i dati. Ad esempio, si consideri il caso di un individuo non enumerato nella sua vera dimora in Somali, ma censito per errore in Oromia. Tale evento non dà luogo ad alcun errore di copertura se i dati sono riferiti a livello nazionale, in quanto l'individuo risulta comunque censito; tuttavia, a livello regionale, il conteggio della popolazione risulta distorto per difetto in Somali (sotto-copertura) e distorto per eccesso in Oromia (sovra-copertura).

demografiche ritenute di particolare interesse.

Per condurre a delle stime finali affidabili, la PES deve soddisfare numerosi vincoli metodologici e operativi (United Nations Secretariat, 2010). In particolare, è fondamentale che l'indagine sia il più possibile indipendente dal Censimento e che, rispetto a questo, venga svolta con maggiore attenzione in ogni sua fase, pur condividendone definizioni e concetti. Tra i vari aspetti ritenuti cruciali per la riuscita della PES rientrano l'impiego sul campo di personale ben addestrato e la possibilità di riconoscere in modo univoco e senza sbagliare le unità statistiche già enumerate dal Censimento.

Anche se, in linea di principio, la PES può essere pianificata in modo tale da fornire stime per entrambe le tipologie di errore (copertura e misura), i requisiti tecnici e i costi di un'indagine di controllo così complessa vanno valutati con attenzione prima di decidere di intraprenderla.

Pertanto, nell'ambito della componente del progetto di cooperazione dedicata alla progettazione della PES, una delle prime attività su cui il team Istat e gli esperti della CSA si sono concentrati ha riguardato la definizione degli obiettivi d'indagine.

Trattandosi della prima esperienza in questo ambito (per i passati Censimenti in Etiopia, la PES non è mai stata realizzata), si è ritenuto opportuno limitare i possibili obiettivi dell'indagine alla:

- valutazione dell'errore di copertura a livello nazionale e, al più, per i principali domini subnazionali, come le regioni e le città amministrative;
- acquisizione di esperienza per migliorare la progettazione e l'implementazione dei futuri Censimenti e delle indagini su larga scala rivolte alle famiglie.

Altri obiettivi – come la valutazione degli errori di misura o la rettifica dei risultati censuari – sono stati ritenuti difficilmente perseguibili con le risorse disponibili nel contesto in esame.

Infine, per semplificare l'indagine senza comprometterne la validità, è stato proposto di circoscrivere la popolazione di interesse della PES alle sole famiglie dimoranti in unità abitative convenzionali – residenziali e/o commerciali – e alle famiglie pastorali, tralasciando altre tipologie residuali prese in considerazione dal Censimento (ad esempio, le convivenze).

5. Disegno di campionamento

Considerata la realtà etiope, anche in relazione alle caratteristiche strutturali della popolazione e alle particolarità della geografia amministrativa (paragrafo 3), per la definizione del disegno di campionamento della PES sono state fornite alla CSA indicazioni circa l'opportunità di:

- considerare la varietà dei domini geografici;
- predisporre la lista di campionamento delle unità territoriali specificando alcune caratteristiche utili alla loro stratificazione quali:
 - tipologia territoriale;
 - dimensione in termini di popolazione;
 - tipologia prevalente di abitazioni;
 - presenza/assenza di rilevanti flussi di popolazione sul territorio;
- preferire l'adozione di un criterio probabilistico per la selezione del campione, rispetto ad un approccio non probabilistico;
- assumere come unità finali di campionamento le sezioni di Censimento (campione areale) e, di conseguenza, includere nel campione finale tutte le famiglie appartenenti alle sezioni estratte.

Queste indicazioni, applicate nel contesto reale, hanno portato alla definizione del disegno di campionamento utilizzato per realizzare la prima indagine pilota sia del Censimento che della PES.

Secondo tale schema, le unità territoriali coincidenti con *wereda* e *kebele* sono state selezionate secondo un disegno di campionamento di tipo areale stratificato a tre stadi in cui, in ogni dominio – le nove regioni e le due città amministrative – i relativi *wereda* sono stati stratificati in tre gruppi (“Urbano”, “Rurale”, “Estremo”).

Il campione finale di sezioni è stato ottenuto secondo il seguente schema:

1° stadio: per ogni dominio, è stato estratto un campione di *wereda* da ogni strato;

- 2° stadio: da ogni *wereda* selezionato nel 1° stadio è stato estratto un campione di *kebele*;
- 3° stadio: tutte le sezioni di Censimento appartenenti ai *kebele* selezionati sono state incluse nel campione finale.

Questo schema si è dimostrato una scelta valida già dalla prima esperienza pilota e pertanto è stato riproposto anche per la PES definitiva.

Riguardo alla dimensione del campione finale, è stato sottolineato agli esperti etiopi che, in assenza di vincoli di costo, questa può dipendere dai seguenti fattori:

- livello di accuratezza delle stime;
- effetto *cluster* – ignoto, dovuto alla correlazione tra gli individui appartenenti alla stessa sezione di Censimento – che incrementa la varianza delle stime;
- piano di diffusione dei risultati finali che definisce i domini di output e i segmenti di popolazione di interesse per la produzione delle stime (ad esempio, per *Sesso*, per classi di *Età*, per *Gruppo etnico*, ecc.).

La dimensione del campione richiesto cresce all'aumentare dell'effetto *cluster*, della precisione desiderata (intervalli di confidenza sufficientemente stretti) e del dettaglio di classificazione delle stime finali. In particolare, se c'è interesse a produrre stime solo a livello nazionale, la dimensione del campione può essere relativamente contenuta rispetto a quella di un campione finalizzato a produrre risultati per domini territoriali più fini.

6. Raccolta dei dati

6.1 Tecniche di rilevazione

Considerato che per l'acquisizione dei dati censuari è prevista la tecnica CAPI (*Computer Assisted Personal Interviewing*) utilizzando dispositivi portatili (*tablet*) dotati di sistema operativo Android¹², la stessa modalità di rilevazione è stata raccomandata anche per la PES.

Come per il Censimento, sono state indicate alla CSA una serie di misure per risolvere eventuali problemi tecnici con i *tablet* durante la fase di raccolta dei dati sul campo, tra le quali anche l'attivazione di centri di assistenza e riparazione.

Rispetto alle tradizionali interviste realizzate con l'utilizzo di questionari cartacei, la tecnica CAPI permette di eliminare la fase di trascrizione delle informazioni dal modello cartaceo al supporto elettronico e di effettuare diversi controlli automatici sui dati al momento della loro immissione, con una conseguente riduzione degli errori e dei tempi di elaborazione. Vale la pena sottolineare che queste peculiarità della tecnica CAPI influiscono positivamente sull'accuratezza dei dati rilevati in entrambe le indagini (Censimento e PES) e, di riflesso, anche su quella dei risultati del *record linkage*.

Comunque, nelle aree "estreme" non fornite da elettricità o con problemi di approvvigionamento dei *tablet*, la tecnica PAPI (*Paper And Pencil Interviewing*) rimane l'unica scelta sia per il Censimento che per la PES.

6.2 Questionario

In generale, il questionario svolge un ruolo centrale nel processo di indagine, poiché è il veicolo attraverso il quale le esigenze informative sono espresse in termini operativi.

Il questionario della PES è stato progettato sulla base del questionario del Censimento al fine di facilitare il confronto delle informazioni utili al riconoscimento delle unità statistiche – famiglie e individui – enumerate in entrambe le indagini. Tipicamente, il questionario della PES è una versione

¹² La scelta strategica di un Censimento *paperless* è stata operata dalla CSA con il supporto dell'UNFPA, sulla base di valutazioni che hanno tenuto conto anche dell'esperienza maturata in altri Paesi.

Per la progettazione del sistema di acquisizione dei dati – disegno del questionario elettronico e sviluppo del sistema di trasmissione dei dati dai *tablet* al *Data Center* – la CSA ha ricevuto assistenza tecnica da USCB e, in particolare, dal team di sviluppo del pacchetto software CSPRO (<https://www.csprouters.org>). Istat e USCB hanno lavorato in stretta collaborazione, sia a distanza sia in loco, per garantire la coerenza degli approcci metodologici proposti alla CSA e la piena integrazione delle componenti software sviluppate con il proprio supporto.

più breve di quello adottato dal Censimento.

Al fine di rilevare quante più informazioni possibili da utilizzare per l’abbinamento, il questionario della PES include gli stessi quesiti – con le stesse opzioni di risposta – previsti nel questionario della prima enumerazione, in relazione a:

1. identificativi della sezione di Censimento;
2. identificativi o caratteristiche socio-demografiche di ciascun individuo.

Con riferimento specifico a quest’ultima categoria di informazioni, si elencano di seguito le variabili rilevate attraverso il questionario relativo all’indagine pilota del Censimento condotta ad agosto 2017, più completo rispetto a quello adottato nell’edizione pilota precedente (novembre 2016):

- *Primo nome;*
- *Secondo nome;*
- *Terzo nome;*
- *Relazione di parentela con il capofamiglia;*
- *Sesso;*
- *Età;*
- *Stato civile;*
- *Gruppo etnico;*
- *Religione;*
- *Titolo di studio.*

Ricalcando l’esperienza positiva dell’Istituto di statistica sudafricano (Statistics South Africa, 2012), è stato proposto agli esperti della CSA l’impiego di un *barcode sticker* – ovvero un codice numerico stampato su supporto adesivo – identificativo della famiglia, in modo da facilitare il riconoscimento univoco delle famiglie e delle persone rilevate in entrambe le occasioni (Censimento e PES).

È stato consigliato di apporre il *barcode sticker* all’ingresso principale di ogni unità abitativa durante le operazioni censuarie, in modo visibile e facilmente individuabile in seguito dai rilevatori della PES. Al riguardo, lo staff della CSA è stato anche sollecitato ad inserire nel manuale dei rilevatori regole precise relative al posizionamento degli adesivi, specificando in ordine di importanza tutti i posti alternativi all’ingresso principale in cui gli *sticker* possono essere attaccati. Inoltre, è stato suggerito che i rilevatori del Censimento riportino nel questionario – o meglio, in un rapporto separato – dove l’adesivo sia stato effettivamente apposto, in modo da facilitare le operazioni finali di riconciliazione.

Con riferimento alle caratteristiche individuali, si è preferito rilevare la variabile *Età* piuttosto che la data di nascita espressa in *Giorno/Mese/Anno*. Infatti, tipicamente in Etiopia quest’ultima informazione presenta maggiori problemi di accuratezza, mancando un sistema di registrazione delle informazioni anagrafiche. Di conseguenza, gli individui – soprattutto quelli meno giovani – possono non conoscere la propria data di nascita con precisione e tendono a riferirla in modo molto approssimato.

Infine, in linea con quanto raccomandato da United Nations Secretariat (2010), il questionario della PES è stato progettato in modo da includere i quesiti utili a classificare ogni individuo come “Non-mover”, “Out-mover”, “In-mover” o “Out-of-scope”¹³, raccogliendo informazioni dettagliate sul suo stato alla data del Censimento.

6.3 Rilevazione sul campo

Come già accennato, per condurre a delle stime finali affidabili, la PES deve soddisfare una serie di vincoli metodologici e operativi. Poiché la validità della metodologia per la stima degli errori di copertura dipende – oltre che dall’ipotesi di abbinamento perfetto delle unità rilevate in entrambe le occasioni – anche dall’assunzione di indipendenza tra il Censimento e la PES (cfr. paragrafo 8), è stata posta particolare attenzione nel tenere il più possibile separate le due indagini.

¹³ “Non-mover”: un individuo rilevato sia alla data del Censimento sia alla data della PES;

“Out-mover”: un individuo rilevato alla data del Censimento, ma non alla data della PES;

“In-mover”: un individuo rilevato alla data della PES, ma non alla data del Censimento;

“Out-of-scope”: un individuo non appartenente alla popolazione di interesse alla data del Censimento (ad esempio, un nato dopo la data del Censimento).

A livello organizzativo, tale obiettivo è stato raggiunto affidando la responsabilità tecnica della PES e delle relative indagini pilota ad una struttura diversa da quella incaricata del Censimento.

Per quanto riguarda le operazioni sul campo, gli esperti della CSA sono stati sensibilizzati, in primo luogo, circa l'importanza di effettuare la raccolta dei dati della PES non oltre un mese di distanza dalle operazioni censuarie, al fine di ridurre al minimo:

- l'impatto dei cambiamenti demografici – nascite, decessi e migrazioni – tra le date di riferimento delle due indagini;
- la possibilità di dichiarazioni falsate dall'effetto memoria.

D'altra parte, questo lasso di tempo è stato ritenuto sufficiente per consentire il recupero di tutto il materiale utilizzato sul campo durante il Censimento, in modo da evitare possibili contaminazioni con le operazioni dell'indagine di controllo.

In secondo luogo, pur utilizzando nella PES rilevatori e supervisori selezionati tra quelli che si riveleranno più validi durante l'enumerazione censuaria, è stato fortemente raccomandato di impiegarli in sezioni diverse da quelle a loro assegnate in precedenza. Inoltre, i rilevatori e i supervisori della PES non dovrebbero conoscere gli esiti dell'enumerazione censuaria nelle sezioni a loro assegnate, in modo da evitare la tendenza a confermare automaticamente quanto già rilevato.

Per ragioni di budget, gli esperti della CSA hanno avanzato l'idea di applicare i *barcode sticker* solo nei *wereda* da cui verranno selezionate le sezioni campione della PES. Tuttavia, occorre sottolineare che questa pratica non è in linea con quanto generalmente raccomandato per lo svolgimento dell'indagine, poiché chi conduce il Censimento non dovrebbe avere alcuna informazione preliminare circa l'indagine di controllo, al fine di preservarne l'indipendenza. Comunque, nel contesto in esame, è stato ritenuto che l'impiego di tali adesivi potesse essere l'unica soluzione per migliorare l'accuratezza del processo di abbinamento (cfr. paragrafo 9).

Infine, per la produzione di reportistica utile al controllo delle operazioni sul campo della PES – ad esempio, statistiche sui questionari restituiti, distribuzione dei dati raccolti, resoconti dei supervisori/rilevatori – è stato proposto alla CSA di avvalersi dell'applicazione Web appositamente sviluppata per monitorare le attività di raccolta dei dati censuari. Tale applicazione, realizzata dall'Istat nell'ambito delle componenti del progetto impegnate a fornire supporto informatico, è stata collaudata con successo in occasione delle varie indagini pilota del Censimento (Bruno *et al.*, 2020).

7. Processo di lavorazione dei dati

7.1 Trattamento preliminare dei dati

Una volta raccolti i dati della PES, occorre abbinarli con quelli rilevati in precedenza nelle stesse sezioni di Censimento. Nel contesto in esame, la procedura di abbinamento comporta il confronto dei campi (o variabili) corrispondenti, comuni ai due insiemi di dati, al fine di riconoscere gli individui enumerati in entrambe le occasioni.

Le variabili utilizzate per l'abbinamento possono essere soggette a errori e/o omissioni e di solito hanno un potere identificativo diverso. Ad esempio, un campo come il *Sesso*, presentando due modalità, contribuisce poco all'individuazione univoca di una persona, a differenza di variabili quali *Nome* o *Età* che, tuttavia, possono essere spesso riferite o trascritte in maniera errata.

Inoltre, bisogna tener conto che nel corso del tempo alcune variabili possono assumere modalità o valori diversi (ad esempio, si pensi allo *Stato civile* di un individuo), rendendo più arduo il riconoscimento di *record* riferibili ad una stessa unità.

Al fine di ridurre il numero di confronti da effettuare, questi possono essere limitati a quei *record* che assumono gli stessi valori per alcune delle variabili di abbinamento che, in questa situazione, sono designate come variabili di "bloccaggio".

Nel contesto in esame, essendo i dati strutturati geograficamente, il codice identificativo della sezione di Censimento può essere considerato una variabile di bloccaggio.

Chiaramente, l'affidabilità e l'efficienza di una procedura di abbinamento dipendono fortemente da come vengono scelte le variabili di bloccaggio, in quanto gli errori in esse contenuti possono

comportare un mancato abbinamento di *record* relativi ad una stessa persona oppure abbinamenti errati.

Pertanto, è stata fortemente raccomandata l'adozione di una serie di misure precauzionali subito dopo l'acquisizione dei dati al fine di migliorare i risultati dell'abbinamento e, di conseguenza, anche l'accuratezza delle stime finali dell'errore di copertura. In particolare, è stato proposto di controllare e processare i *record* provenienti dalle due indagini, Censimento e PES, in modo da:

- evidenziare e risolvere eventuali errori relativi ai codici identificativi delle sezioni di Censimento;
- verificare l'esistenza e l'unicità del codice identificativo di ciascuna famiglia;
- rilevare e rimuovere eventuali *record* duplicati;
- eseguire la standardizzazione delle informazioni in formato libero (nomi degli individui).

Per quanto riguarda la necessità di standardizzare i nomi degli individui, vale la pena notare che queste informazioni possono essere riferite o riportate in modo diverso in entrambe le indagini (Censimento e PES). Infatti, spesso un nominativo, pur se relativo ad uno stesso soggetto, può non coincidere a causa della presenza di parole non significative, sigle ed abbreviazioni¹⁴, banali errori di trascrizione o modi alternativi utilizzati per denominare la stessa persona. Inoltre, nel contesto in esame, differenze rilevanti nella scrittura di uno stesso nome possono derivare anche dalla traduzione dall'alfabeto amarico a quello latino. Situazioni di questo tipo potrebbero comportare un numero considerevole di mancati abbinamenti.

Pertanto, la standardizzazione dei nomi dovrebbe essere effettuata rimuovendo i caratteri più soggetti a errori, tenendo conto delle caratteristiche fonetiche delle lingue parlate nel paese.

A questo proposito, si descrive di seguito un algoritmo utilizzabile per la standardizzazione dei nomi, messo a punto e testato con successo in occasione delle prime due esperienze pilota realizzate durante l'attuazione del progetto di cooperazione.

L'algoritmo produce una forma standardizzata di quattro caratteri per ciascuna componente del nome di ogni individuo (*Primo nome*, *Secondo nome*, *Terzo nome*):

- Passo 1* eventuali titoli, parole di supporto, punteggiatura, spazi vuoti e caratteri speciali vengono rimossi dalla componente del nome;
- Passo 2* la lettera iniziale – vocale o consonante – della componente risultante dal *Passo 1* diventa la prima lettera della versione standardizzata; tuttavia
- se la lettera iniziale è “I” o “E”, questa è sostituita da “1”;
 - se la lettera iniziale è “O” o “U”, questa è sostituita da “2”;
- Passo 3* le successive due consonanti e l'ultima consonante della componente risultante dal *Passo 1* sono inserite nella forma standardizzata nello stesso ordine in cui compaiono; tuttavia, le consonanti ripetute e adiacenti nel nome originale sono considerate una sola volta;
- Passo 4* se il numero totale di cifre e/o lettere della forma standardizzata risultante al *Passo 3* è inferiore a quattro, questa viene completata da tutte le “X” necessarie, fino al raggiungimento di quattro caratteri di lunghezza.

7.2 Abbinamento dei dati

Dal momento che errori di abbinamento, anche molto contenuti, possono dar luogo a distorsioni di entità non trascurabile nelle stime finali, occorre prestare particolare attenzione all'accuratezza di tutto il processo di integrazione dei dati. In tale ambito, gli errori possono essere di due tipi:

- I. *record* effettivamente relativi ad uno stesso individuo non vengono abbinati;
- II. *record* relativi a individui diversi vengono comunque abbinati.

Se si richiede la coincidenza perfetta dei valori di tutte le variabili di abbinamento nei due insiemi di dati per stabilire che uno stesso individuo è stato enumerato in entrambe le occasioni, è inevitabile

¹⁴ L'elenco delle abbreviazioni o parole che possono essere trascurate nel contesto specifico è lungo e comprende, ad esempio, H. (che sta per Hayile) e G. (che sta per Gebre o Gebru). In particolare, Gebre è un comune nome etiopico maschile – che significa “servo” – usato spesso in alcuni nomi composti come Gebre Meskel (“Servo della croce”) o Gebre Mariam (“Servo di Maria”), mentre Gebru è una variante tipica della regione del Tigray.

che parecchi *record* risultino erroneamente non abbinati. Ciò può accadere per una serie di motivi, tra cui il verificarsi di:

- mancate risposte;
- errori nella fase di rilevazione dei dati;
- variazioni durante il periodo di tempo trascorso tra le due rilevazioni;
- errori nel trattamento preliminare che i dati subiscono prima di essere sottoposti all'abbinamento.

Tuttavia, se si rende più flessibile la regola di decisione – ad esempio, ammettendo la non coincidenza di qualcuno dei valori delle variabili di abbinamento – in modo da diminuire i casi di errore di tipo I, possono aumentare i casi di errore di tipo II.

Per integrare i dati del Censimento con quelli della PES è stata proposta una strategia articolata in più fasi, in ognuna delle quali un'operazione automatica di abbinamento deterministico precede un trattamento manuale assistito da computer. L'insieme delle variabili di abbinamento (chiamato anche chiave di abbinamento) utilizzato nella parte automatica di ogni fase è stato definito tenendo conto del diverso potere identificativo e dell'accuratezza delle variabili disponibili (cfr. sotto-paragrafo 6.2).

In effetti, un processo di abbinamento dei dati completamente manuale sarebbe troppo costoso e richiederebbe molto tempo a causa della numerosità dei due insiemi di dati prodotti dal Censimento e dalla PES. È apparso prioritario automatizzare, quanto più possibile, le varie operazioni; il ricorso all'abbinamento manuale, condotto comunque con l'ausilio di interfacce opportunamente predisposte, viene limitato ai casi in cui ciò è strettamente necessario, con notevole guadagno in termini di accuratezza e riproducibilità dei risultati.

Tale strategia è stata messa in pratica per elaborare i dati delle prime due esperienze pilota realizzate durante il progetto di cooperazione. In particolare, è stata sviluppata l'applicazione Web prototipale *CSPro Linkage* per eseguire/assistere l'abbinamento dei dati della PES con i dati censuari; tale applicazione – che utilizza Java, Spring MVC Framework e procedure MySQL – effettua anche la standardizzazione dei nomi degli individui secondo l'algoritmo descritto nel sotto-paragrafo 7.1.

Alcune linee guida per l'uso dell'applicazione *CSPro Linkage* sono riportate in appendice.

Al suo stato attuale di sviluppo, l'applicazione elabora i dati elementari del Censimento e della PES in undici fasi, considerando le sezioni di Censimento come blocchi distinti.

Ogni fase è costituita da un passo iniziale di *record linkage* di tipo deterministico che produce abbinamenti ogni qualvolta il valore della chiave utilizzata:

- non contenga alcuna mancata risposta parziale;
- identifichi univocamente i *record*;
- coincida perfettamente nelle due occasioni, Censimento e PES.

Dopo questa operazione completamente automatica, segue un passo di abbinamento manuale (assistito da computer) che utilizza tutte le informazioni disponibili e riguarda gli eventuali familiari – non abbinati – degli individui agganciati in modo automatico nell'ambito della stessa fase. L'unica differenza tra le varie fasi è nell'insieme di variabili utilizzate per l'abbinamento automatico:

- Fase 0* tutte le variabili comuni (*Numero del barcode sticker, Primo nome, Secondo nome, Terzo nome, Relazione di parentela con il capofamiglia, Sesso, Età, Stato civile, Titolo di studio, Religione, Gruppo etnico*);
- Fase 1* tutte le variabili comuni ad eccezione di *Età*;
- Fase 2* componenti del nome in forma standardizzata (*Primo nome standardizzato, Secondo nome standardizzato, Terzo nome standardizzato*) e tutte le altre variabili comuni (*Numero del barcode sticker, Relazione di parentela con il capofamiglia, Sesso, Età, Stato civile, Titolo di studio, Religione, Gruppo etnico*);
- Fase 3* componenti del nome in forma standardizzata e tutte le altre variabili comuni ad eccezione di *Età*;
- Fase 4* componenti del nome in forma standardizzata, *Numero del barcode sticker* e *Sesso*;
- Fase 5* prime due componenti del nome in forma standardizzata (*Primo nome standardizzato, Secondo nome standardizzato*), *Numero del barcode sticker* e *Sesso*;

- Fase 6* tutte le variabili comuni ad eccezione di *Numero del barcode sticker* e *Età*;
- Fase 7* componenti del nome in forma standardizzata e tutte le altre variabili comuni ad eccezione di *Numero del barcode sticker*;
- Fase 8* componenti del nome in forma standardizzata e tutte le altre variabili comuni ad eccezione di *Numero del barcode sticker* e *Età*;
- Fase 9* componenti del nome in forma standardizzata e *Sesso*;
- Fase 10* prime due componenti del nome in forma standardizzata e *Sesso*.

Ulteriori sforzi nello sviluppo dell'applicazione dovrebbero essere fatti per estendere la ricerca delle informazioni da abbinare alle sezioni di Censimento confinanti o anche a blocchi più ampi delle sezioni, come quelli individuati dalle unità amministrative *kebele* e *woreda*.

Ciò può essere utile quando una persona o una famiglia non viene trovata in una determinata sezione di Censimento (potrebbero esserci dei casi enumerati al Censimento in una determinata sezione, ma rilevati successivamente dalla PES come non dimoranti in quella sezione, e viceversa).

Comunque, è stato raccomandato agli esperti della CSA di prevedere delle visite di riconciliazione sul campo in modo da raccogliere informazioni che permettano di fare maggiore chiarezza sui casi più incerti e risolvere le ambiguità risultanti dal processo di abbinamento.

8. Procedimento di stima

Con riferimento al contesto etiope, trattandosi della prima esperienza sulla PES per la CSA (cfr. paragrafo 4), l'obiettivo primario dell'indagine è stimare il tasso di sotto-copertura, definito come rapporto tra il numero di persone non conteggiate al Censimento e il dato (stimato) di popolazione.

La metodologia che è stata proposta agli esperti della CSA per la stima di tale tasso è quella *Dual System*, basata sulla tecnica di cattura-ricattura. Questa metodologia permette di stimare la numerosità di una popolazione sconosciuta attraverso due enumerazioni delle unità eleggibili e il successivo riconoscimento dei casi rilevati nelle due occasioni, utilizzando tecniche di *record linkage* (Wolter, 1986).

Con riferimento alle sezioni di Censimento estratte per il campione della PES, i conteggi risultanti possono essere rappresentati nella seguente tabella:

	RILEVATI ALLA PES	NON RILEVATI ALLA PES	TOTALE
CENSITI	x_{11}	x_{12}	x_1
NON CENSITI	x_{21}		
TOTALE	$x_{.1}$		N

dove x_1 è il numero di individui conteggiati al Censimento, x_{11} è il numero di individui enumerati sia al Censimento che alla PES (risultato del processo di abbinamento dei dati) e N rappresenta l'ammontare totale di popolazione (valore ignoto).

Sotto l'ipotesi che le due indagini siano indipendenti l'una dall'altra – cioè la probabilità che un'unità sia rilevata alla PES non dipende dal fatto che l'unità sia stata censita o meno – e supponendo che non vi sia sovra-copertura nel Censimento, la stima della popolazione totale è data da:

$$\hat{N} = \frac{x_1 \cdot x_{.1}}{x_{11}}$$

La validità del metodo di stima proposto, oltre che dal presupposto dell'indipendenza tra i due tipi di conteggio, dipende anche dalla validità delle seguenti ipotesi di base:

- la popolazione non varia tra le due indagini;
- il numero di unità abbinate è determinato senza errori (accuratezza dei risultati del *record linkage*);
- in ogni indagine, tutte le unità hanno la stessa probabilità di essere enumerate.

Considerando la limitata esperienza in tema di PES nella CSA e la mancanza di informazioni

ausiliarie da utilizzare nel processo di stima, è stato suggerito di adottare metodi di stima diretta, eventualmente impiegando un procedimento di calibrazione.

I tassi di copertura del Censimento e di sotto-copertura possono essere stimati come segue:

$$\text{tasso di copertura: } \frac{\hat{x}_{11}}{\hat{x}_{.1}} = \frac{x_1}{\widehat{N}}$$

$$\text{tasso di sotto-copertura: } 1 - \frac{\hat{x}_{11}}{\hat{x}_{.1}} = 1 - \frac{x_1}{\widehat{N}}$$

L'adozione dello stimatore calibrato richiede il calcolo di due fattori:

- i. il peso diretto o di campionamento, indotto dal disegno campionario adottato, pari all'inverso della probabilità di selezione delle unità finali (le sezioni di Censimento), applicato a tutti gli individui appartenenti alla stessa sezione;
- ii. il fattore di riponderazione (calibrazione), che permette di calibrare il peso diretto individuale al fine di assicurare la coerenza tra alcuni totali di popolazione ottenuti al Censimento (ad esempio, la popolazione per *Sesso*, per classi di *Età*, per *Gruppo etnico*) e le relative stime, calcolate utilizzando i dati della PES rilevati nelle sezioni campione.

Se si utilizzano solo i pesi diretti, lo stimatore della sotto-copertura è:

$$1 - \frac{\sum_{j \in S} x_{11,j} w_j}{\sum_{j \in S} x_{.1,j} w_j}$$

dove:

$x_{11,j}$ è il numero di unità abbinata, appartenenti alla sezione j ;

$x_{.1,j}$ è il numero di unità enumerate alla PES, appartenenti alla sezione j ;

w_j è il peso diretto associato a tutte le unità appartenenti alla sezione j ;

S è l'insieme delle sezioni campione della PES.

Se si usano i pesi calibrati, lo stimatore della sotto-copertura diventa:

$$1 - \frac{\sum_{i \in U} z_{11,i} w_i^*}{\sum_{i \in U} z_{.1,i} w_i^*}$$

dove:

$z_{11,i}$ è l'indicatore di abbinamento riferito all'unità i (1 se l'unità è stata abbinata, 0 altrimenti);

$z_{.1,i}$ è l'indicatore di eleggibilità riferito all'unità i (1 se l'unità è eleggibile per la PES, 0 altrimenti);

w_i^* è il peso calibrato (pari al prodotto tra il peso diretto e il fattore di calibrazione) per l'unità i ;

U è l'insieme delle unità appartenenti alle sezioni campione della PES.

Nel contesto in esame, il tasso di sotto-copertura può essere prodotto a livello nazionale e per le undici regioni/città amministrative.

Dato il disegno di campionamento che prevede la stratificazione dei *wereda*, i tassi di sotto-copertura possono essere calcolati anche per i tre tipi di *wereda* ("Urbano", "Rurale", "Estremo"). Tuttavia, è importante ricordare che la possibilità di diffondere risultati finali accurati riferiti a domini più fini – sotto-regioni o particolari sotto-popolazioni identificate da specifiche caratteristiche socio-demografiche (ad esempio, *Sesso*, classe di *Età*, *Gruppo etnico*) – dipende dalla dimensione e dall'allocazione del campione utilizzato.

9. Esperienze e risultati acquisiti dalle prime due indagini pilota

Durante il periodo di attuazione del progetto è stato possibile valutare i risultati di due indagini pilota condotte sia per il Censimento sia per la PES, su uno stesso insieme di sezioni campione.

Per quanto riguarda la prima esperienza pilota, questo insieme è risultato costituito da 38 sezioni¹⁵ campione, selezionate a partire dalle 256 sulle quali è stata condotta la prima prova censuaria (cfr. paragrafo 5).

In questa esperienza pilota, che ha visto la cattura dei dati svolgersi a tre mesi di distanza (novembre 2016, per il Censimento; febbraio 2017, per la PES), sono state osservate differenze molto marcate nei totali degli individui enumerati nelle due occasioni a livello di sezione di Censimento.

Tali differenze sono state ricondotte a problemi nella trasmissione dei dati, mancato impiego di intervistatori addestrati ed esperti, ampio lasso di tempo intercorso tra le due indagini pilota (Censimento e PES). Ciò ha portato inevitabilmente a un numero considerevole di mancati abbinamenti, inficiando di fatto l'esito finale dell'esperimento.

Questo primo ciclo di indagini pilota è stato utile ad evidenziare le principali problematiche tecniche connesse alla conduzione di una rilevazione censuaria in Etiopia, in particolar modo quelle relative a:

- sostenibilità della raccolta dei dati mediante strumenti informatici (problemi infrastrutturali);
- professionalità dei rilevatori (problemi della rete di rilevazione);
- tempistica delle operazioni sul campo (problemi organizzativi).

Alcuni progressi in tal senso sono stati compiuti in occasione della seconda esperienza pilota. In questo secondo ciclo di indagini i dati della PES sono stati rilevati (con tecnica CAPI) su un insieme di quattro sezioni campione selezionate in modo ragionato a partire dalle dieci sulle quali è stato condotto l'esercizio censuario corrispondente.

Questa volta, lo svolgimento delle indagini a solo un mese di distanza (agosto 2017, per il Censimento; settembre 2017, per la PES), il superamento di alcuni problemi nella trasmissione dei dati, l'utilizzo di intervistatori meglio addestrati e l'introduzione del *barcode sticker* hanno portato a una maggiore accuratezza dei dati raccolti.

Il secondo ciclo di indagini, fatto tesoro dell'esperienza pilota precedente, ha permesso di far emergere problematiche più sottili, legate al riconoscimento univoco degli individui rilevati nelle due occasioni. Infatti, nonostante i progressi compiuti, soltanto per una delle quattro sezioni di Censimento incluse nel campione della PES i risultati dell'abbinamento sono stati ritenuti soddisfacenti (il 93 per cento degli individui rilevati al Censimento sono stati abbinati); per le altre tre sezioni, la percentuale di individui abbinati è risultata non superiore all'80 per cento¹⁶.

In entrambe le esperienze pilota, la criticità principale è stata rappresentata dal fatto che il tipo di informazione acquisita attraverso il questionario del Censimento – e quindi della PES – si è rivelato non abbastanza discriminante, ovvero non ha consentito di riconoscere gli individui enumerati nelle due occasioni in modo univoco. Infatti, all'interno di una stessa sezione di Censimento si sono verificati casi in cui più di una persona presentava gli stessi valori per tutte le variabili di abbinamento. In particolare, con riferimento ai nomi degli individui, la frequenza con cui si sono ripetute le varie componenti – *Primo nome*, *Secondo nome*, *Terzo nome* – è stata tale da ridurre considerevolmente il potere discriminante dei nomi¹⁷.

Alcuni esempi illustrativi basati sui dati della seconda indagine pilota della PES sono riportati nelle Tavole 1 e 2.

¹⁵ Su 30 sezioni la rilevazione dei dati è stata effettuata con tecnica CAPI, mentre sulle restanti 8 è stata utilizzata la modalità PAPI.

¹⁶ Nei Paesi in cui il Censimento della Popolazione è caratterizzato da un basso errore di copertura, la percentuale di individui abbinati risulta generalmente molto elevata (superiore al 98 per cento).

¹⁷ A questo proposito, vale la pena osservare che in Etiopia è molto diffusa l'usanza di attribuire nomi che coincidono con il *Primo nome* di parenti prossimi (padre, madre, nonno/a, ecc.).

Tavola 1. Individui rilevati in una stessa sezione di Censimento che presentavano lo stesso valore di Primo nome standardizzato e Secondo nome standardizzato

ID	PRIMO NOME	SECONDO NOME	TERZO NOME	PRIMO NOME STANDARDIZZATO	SECONDO NOME STANDARDIZZATO
2383	Zehara	Hamed	Bilai	ZHRX	HMDX
2650	Zehara	Hamed	Niebale	ZHRX	HMDX
2818	Zehara	Hamed	Ali	ZHRX	HMDX
3577	Zehara	Hamedu	Yaldu	ZHRX	HMDX
3774	Zehara	Hamed	Dimio	ZHRX	HMDX
3805	Zehara	Hamed	Ebrahim	ZHRX	HMDX

Fonte: Seconda indagine pilota della PES (settembre 2017), CSA

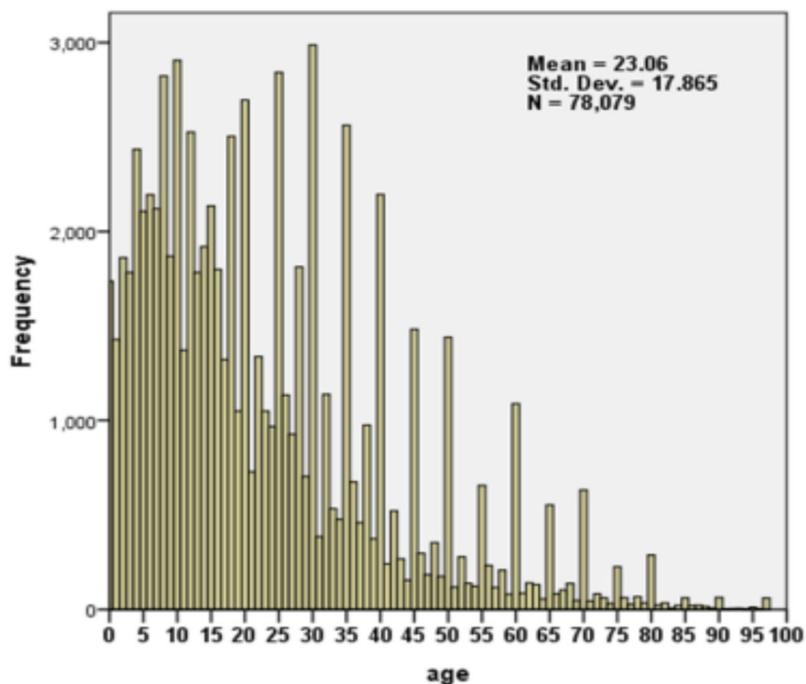
Tavola 2. Valori molto frequenti di Primo nome standardizzato, Secondo nome standardizzato, Terzo nome standardizzato osservati sulle quattro sezioni di Censimento

VARIABILE	FREQUENZA
PRIMO NOME STANDARDIZZATO	
MHMD	61
FTMX	43
HSNX	35
SECONDO NOME STANDARDIZZATO	
MHMD	111
HSNX	76
ALXX	75
TERZO NOME STANDARDIZZATO	
MHMD	110
ALXX	84
HSNX	66

Fonte: Seconda indagine pilota della PES (settembre 2017), CSA

Un discorso a parte va fatto per la variabile *Età*. Come già accennato (cfr. sotto-paragrafo 6.2), poiché non esiste un registro della popolazione in Etiopia, la maggior parte delle persone non conosce o non ricorda la propria data di nascita. Pertanto, non sorprende che la distribuzione di frequenza per *Età* degli individui evidenzia una chiara tendenza a dichiarare l'età utilizzando multipli di 5.

Il grafico basato su dati della prima indagine pilota del Censimento mostra che il fenomeno dell'arrotondamento è più accentuato per la popolazione adulta (Figura 2). Ovviamente, ciò riduce sia l'accuratezza sia il potere identificativo della variabile *Età* nel processo di riconoscimento degli individui.

Figura 2. Distribuzione di frequenza per Età degli individui

Fonte: Prima indagine pilota del Censimento (novembre 2016), CSA

Con specifico riferimento all'obiettivo primario della PES (cfr. paragrafo 8), i risultati della prima pilota non sono stati utilizzati per il calcolo dei tassi di sotto-copertura, a causa dell'elevato numero di errori di abbinamento che hanno compromesso l'esito dell'esperimento.

La seconda pilota della PES, invece, pur tenendo conto delle problematiche accennate in precedenza e dei limiti legati all'esiguità del numero di sezioni campione coinvolte, ha permesso la stima di un tasso di sotto-copertura. In questo caso, la stima del tasso complessivo di sotto-copertura ottenuta applicando diversi sistemi di ponderazione varia dal 15,94 al 22,41 per cento, a seconda che si adottino un sistema di pesi più o meno favorevole, intendendo per favorevole un sistema di ponderazione che attribuisce un peso maggiore alle sezioni di Censimento in cui la percentuale di abbinamento è più elevata.

10. Considerazioni conclusive

Il supporto all'Etiopia si è inserito nel quadro delle complesse e articolate attività di preparazione del 4° Censimento della Popolazione e delle Abitazioni, cui hanno contribuito, con differenti modalità, diversi donatori e partner internazionali nell'impegno per lo sviluppo.

Specificamente, il progetto italiano ha seguito due direttrici principali, l'una finalizzata alla predisposizione dell'infrastruttura informatica per il Censimento e l'altra concentrata prevalentemente sulla progettazione della PES e sullo sviluppo di strumenti informatici per il *record linkage*.

Con riferimento a quest'ultima direttrice, durante il periodo di attuazione del progetto è stato possibile valutare la bontà delle principali soluzioni metodologiche ed operative proposte dal team Istat per la PES, sulla base di quanto emerso dalle varie indagini pilota condotte dalla CSA.

In particolare, è stata verificata l'opportunità di alcune delle scelte operate per il disegno di campionamento e per le fasi di raccolta e lavorazione dei dati, traendo una serie di conclusioni che sono state riportate al partner etiope. Tali conclusioni si basano sui risultati relativi alle indagini effettivamente concluse alla data di scadenza del progetto (giugno 2018).

Per ciò che concerne il disegno di campionamento, lo schema areale stratificato a tre stadi si è dimostrato una scelta valida già dalla prima esperienza pilota e pertanto è stato riproposto anche per la PES definitiva. Inoltre, è stata sottolineata agli esperti della CSA l'importanza di determinare la dimensione del campione finale tenendo conto di fattori quali il livello di accuratezza e il dettaglio di classificazione desiderati per le stime finali.

Con riferimento alla fase di raccolta dei dati, è stato confermato quanto già impostato per il Censimento, raccomandando l'adozione in via preferenziale della tecnica CAPI e circoscrivendo l'uso della modalità PAPI solo alle aree "estreme" non fornite da elettricità o con problemi di approvvigionamento dei *tablet*. Inoltre, per la produzione dei vari tipi di reportistica utili al controllo delle operazioni sul campo della PES, è stato suggerito di avvalersi dell'applicazione Web, appositamente sviluppata per monitorare le attività di rilevazione sul campo dei dati censuari e collaudata con successo in occasione delle varie indagini pilota del Censimento.

Nel corso delle varie missioni, il team Istat ha sempre cercato di sensibilizzare gli esperti della CSA circa la necessità di porre particolare attenzione nella raccolta delle informazioni da utilizzare al fine di riconoscere le unità statistiche enumerate sia al Censimento sia alla PES, nonché nell'adozione delle soluzioni utili a garantire il rispetto delle assunzioni alla base della metodologia di stima dell'errore di copertura. Al riguardo, vale la pena sottolineare come alcune misure introdotte in occasione della seconda esperienza pilota – utilizzo del *barcode sticker*, svolgimento delle indagini a solo un mese di distanza, impiego di intervistatori ben addestrati – abbiano permesso di compiere progressi apprezzabili in termini di accuratezza dei dati raccolti, rispetto al primo esercizio dai risultati deludenti.

Per l'integrazione dei dati censuari con quelli della PES, è stato proposto di utilizzare, con opportune migliorie, l'applicazione Web prototipale già impiegata in occasione delle esperienze pilota per effettuare la standardizzazione dei nomi degli individui ed eseguire/assistere l'abbinamento dei *record*. Una miglioria auspicabile per l'applicazione riguarda la possibilità di effettuare la ricerca delle informazioni da integrare anche su blocchi più ampi delle sezioni di Censimento.

Per quanto riguarda la procedura di calcolo del tasso di copertura, vista la limitata esperienza in tema di PES nella CSA e la mancanza di informazioni ausiliarie, è stato suggerito di adottare metodi di stima diretta, impiegando eventualmente un procedimento di calibrazione.

Alla luce dei risultati emersi dalle due esperienze pilota – e, specificamente, delle valutazioni sulla copertura ottenute ipotizzando sistemi di ponderazione corrispondenti a scenari più o meno favorevoli – si è provveduto ad allertare il personale della CSA circa le criticità esistenti.

In particolare, è stato evidenziato come le questioni strutturali che caratterizzano il contesto etiopico – in particolare, quelle derivanti dalla mobilità della popolazione pastorale e dalla debolezza delle informazioni utilizzabili per l'abbinamento – possano creare problemi sia nella fase di rilevazione che nel processo di lavorazione dei dati, introducendo un effetto distorsivo non trascurabile nelle stime finali dei tassi di copertura.

Al fine di mitigare tale effetto, la CSA dovrebbe compiere ulteriori sforzi – non facilmente attuabili, almeno nel breve periodo – per migliorare l'accuratezza di ciascuna fase della PES, dalla progettazione del campione alla predisposizione delle procedure di stima.

Pertanto, è stata consigliata la massima cautela nella diffusione dei risultati finali dell'indagine di controllo della copertura, nel caso in cui venisse effettivamente condotta. I risultati dovrebbero contribuire, piuttosto, a leggere e interpretare in modo più consapevole i dati del Censimento e a migliorare la progettazione e la realizzazione delle indagini future di vasta portata sulle famiglie.

Appendice

Requisiti di sistema

L'applicazione Web prototipale *CSPro Linkage* esegue/assiste l'abbinamento dei dati della PES con i dati censuari per una sezione di Censimento alla volta. I microdati sono elaborati in diversi passaggi, verificando la corrispondenza tra le informazioni rilevate nelle due indagini, secondo la strategia descritta nel sotto-paragrafo 6.2.

L'applicazione è stata sviluppata utilizzando Java, Spring MVC Framework e procedure MySQL. Per utilizzarla, è necessaria l'installazione dei seguenti prodotti software:

- JAVA 1.7+;
- MYSQL 5.7.

Inoltre, è richiesta la seguente libreria:

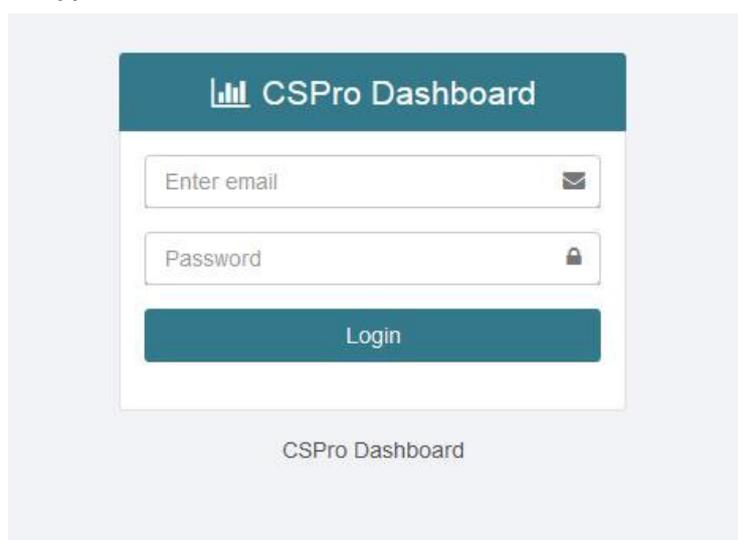
- MySQL Connector/J 5.1.40 (mysql-connector-java-5.1.40-bin.jar).

Il progetto *CSPro Linkage NetBeans Java* è disponibile su *github*:

<https://github.com/IstatCooperation/CSProLinkage>.

L'applicazione può essere utilizzata nelle due modalità "Amministratore" e "Operatore", inserendo le opportune credenziali (Figura 3).

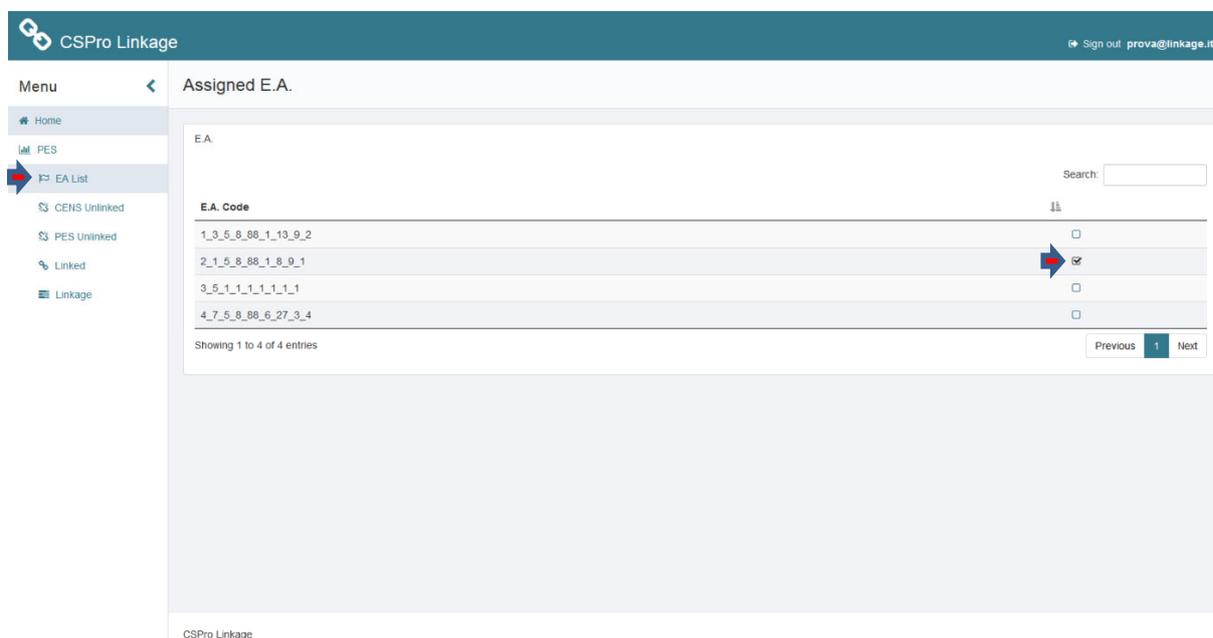
Figura 3. Accesso all'applicazione



Nel prosieguo vengono illustrate le principali funzionalità dell'applicazione per entrambe le modalità, utilizzando dati non reali.

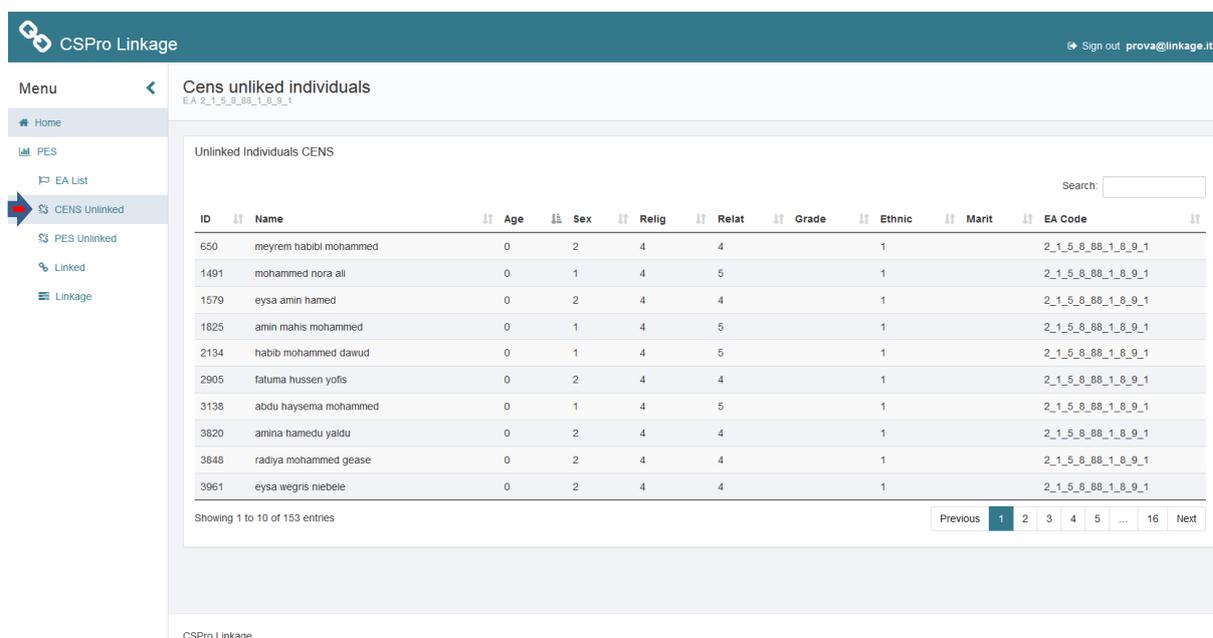
Interfaccia operatore

Figura 4. Selezione di una sezione di Censimento



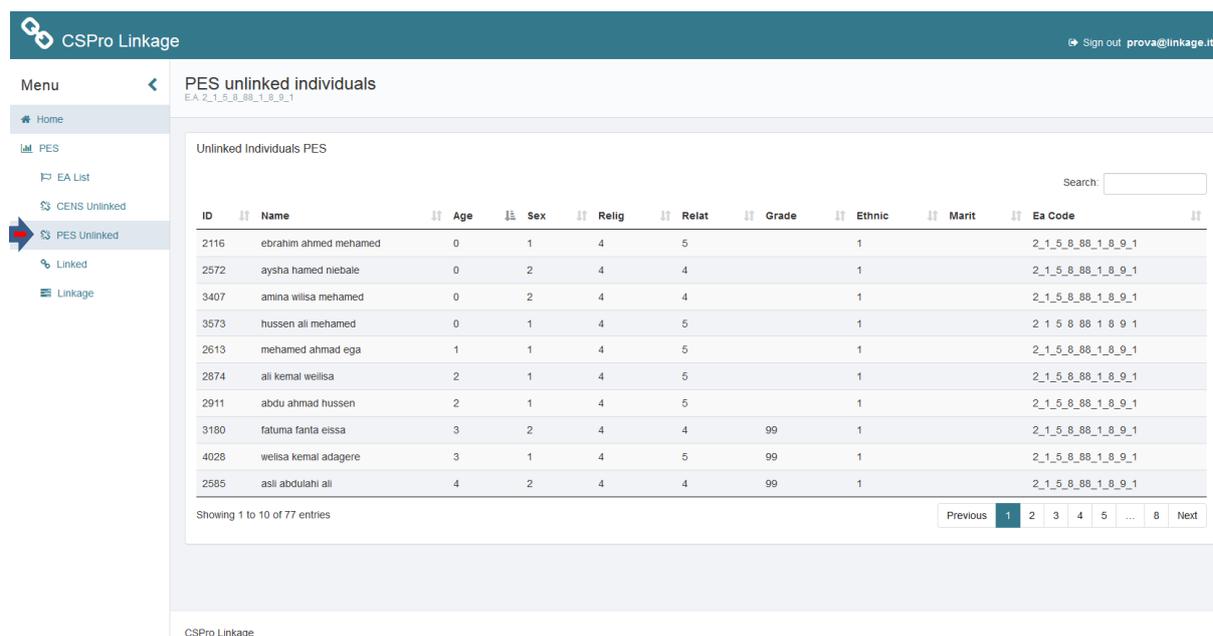
Cliccando su “EA List” (Figura 4), è possibile visualizzare la lista delle sezioni campione della PES, assegnate all’utente connesso. La colonna “EA Code” riporta il codice identificativo della sezione di Censimento. La sezione prescelta può essere selezionata spuntando la casella corrispondente.

Figura 5. Esplorazione delle informazioni relative agli individui censiti e non abbinati



Cliccando su “CENS Unlinked” (Figura 5), è possibile visualizzare la lista degli individui censiti e non abbinati.

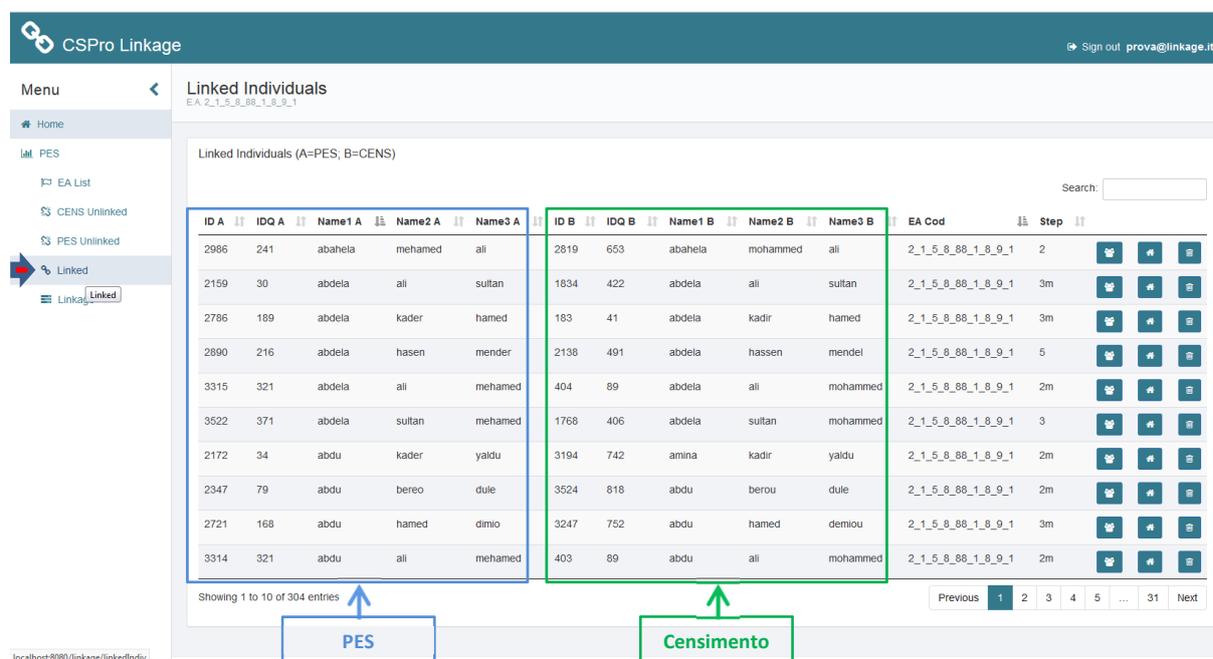
Figura 6. Esplorazione delle informazioni relative agli individui enumerati alla PES e non abbinati



Cliccando su “PES Unlinked” (Figura 6), è possibile visualizzare la lista degli individui enumerati alla PES e non abbinati.

Cliccando su “Linked” (Figura 7), viene mostrata la lista di tutti gli abbinamenti effettuati (in modo automatico o manuale).

Figura 7. Esplorazione delle informazioni relative agli individui abbinati



Per ogni individuo abbinato (Figura 7), sono riportate alcune delle informazioni rilevate sia alla PES (etichettate con “A”) che al Censimento (etichettate con “B”). La colonna “Step” indica in quale passo dell’applicazione è stato effettuato l’abbinamento (ad esempio, “3m” significa che l’abbinamento è stato eseguito manualmente dopo il passo automatico della fase 3).

Figura 8. Esplorazione delle informazioni relative ai familiari (non abbinati) di individui abbinati automaticamente

ID A || IDQ A || Name1 A || Name2 A || Name3 A || ID B || IDQ B || Name1 B || Name2 B || Name3 B || EA Cod || Step ||

7796 1926 abadit gesesew getahun 11571 2878 Abadit Gesesew Getahun 1-4-4-1-1-1-1-5-2 0

Families

HH Quest PES 1926

ID	Name	Age	Sex	Relig	Relat	Grade	Tongue	Marit
7802	goytu hayle gesesew	6	1	1	9	99	80	<input checked="" type="radio"/>
7796	kiros gebre gerase	55	2	1	1	99	80	6
7801	tegsti hayle gesesew	9	2	1	9	2	80	

Showing 1 to 3 of 3 entries

HH Quest CENS 2878

ID	Name	Age	Sex	Relig	Relat	Grade	Tongue	Marit
11574	Goitom Haile Gesesew	6	1	1	9		80	<input checked="" type="radio"/>
11568	Letekiros Gebre Gerase	59	2	1	1		80	6
11573	Tigist Haile Gesesew	8	2	1	9	2	80	

Showing 1 to 3 of 3 entries

Link Individuals

Cliccando sul primo tasto nella riga corrispondente a un individuo abbinato automaticamente (Figura 8), possono essere visualizzate le informazioni relative ai suoi familiari non ancora abbinati (questo tasto è presente in entrambe le pagine “Linked” e “Linkage” ad ogni fase dell’applicazione).

A questo punto, si apriranno incolonnati i due elenchi di familiari non abbinati, enumerati alla PES e al Censimento. I *radio button* consentono di selezionare manualmente i *record* da collegare; cliccando su “Link Individuals”, è poi possibile abbinare i *record* selezionati.

Figura 9. Esplorazione di informazioni relative a familiari abbinati

ID A	IDQ A	Name1 A	Name2 A	Name3 A	ID B	IDQ B	Name1 B	Name2 B	Name3 B	EA Cod	Step
7798	1926	abadit	gesesew	getahun	11571	2878	Abadit	Gesesew	Getahun	1-4-4-1-1-1-1-5-2	0
4837	1204	abebe	gezael	desalegn	11248	2791	Abeba	Gezae	Desalegn	1-4-4-1-1-1-1-5-2	2

Linked Individuals

HH Quest PES 1204
HH Quest CENS 2791

ID A	Name A	Age A	Sex A	ID B	Name B	Age B	Sex B	Step
4837	abebe gezael desalegn	10	2	11248	Abeba Gezae Desalegn	10	2	2
4835	gezael desalegn alemu	37	1	11246	Gezae Desalegn Alemu	39	1	4
4839	zewdu gezael desalegn	4	2	11250	Zewdu Gezae Desalegn	4	2	4

Showing 1 to 3 of 3 entries

Close

Cliccando sul simbolo della casa (Figura 9), per l'individuo abbinato della riga corrispondente è possibile visualizzare le due liste dei familiari abbinati, enumerati sia alla PES che al Censimento (questo tasto è presente in entrambe le pagine **“Linked”** e **“Linkage”** ad ogni fase dell'applicazione).

Figura 10. Rimozione di un abbinamento

ID A	IDQ A	Name1 A	Name2 A	Name3 A	ID B	IDQ B	Name1 B	Name2 B	Name3 B	EA Cod	Step
7798	1926	abadit	gesesew	getahun	11571	2878	Abadit	Gesesew	Getahun	1-4-4-1-1-1-1-5-2	0
4837	1204	abebe	gezael	desalegn	11248	2791	Abeba	Gezae	Desalegn	1-4-4-1-1-1-1-5-2	2

Remove link

Remove link between

Individual A 4837
Individual B 11248

➡
Remove link
Close

Nella pagina **“Linked”** (Figura 10), è possibile rimuovere un abbinamento cliccando sul simbolo del cestino della riga corrispondente (questo tasto è disponibile per ogni fase dell'applicazione, tranne che per la fase 0).

A questo punto, si aprirà un *pop-up* per confermare o meno la scelta. Cliccando su **“Remove link”**, è quindi possibile rimuovere l'abbinamento selezionato.

Figura 11. Esecuzione di un passo di abbinamento

The screenshot shows the 'Step 2' interface of the CSPro Linkage application. The title bar reads 'Step 2' with the identifier 'EA_2_1_5_8_88_1_8_9_1'. The sidebar menu on the left has 'Linkage' highlighted. The main content area displays a table of linked individuals. The table has the following columns: ID A, Name A, ID B, Name B, Age, Sex, Religion, Relat, Grade, Ethnic, Marital, and Cod EA. The table contains 10 rows of data, each with a '0.00' value and a star icon in the right margin. Below the table, it says 'Showing 1 to 10 of 20 entries' and includes 'Previous', '1', '2', and 'Next' navigation buttons.

ID A	Name A	ID B	Name B	Age	Sex	Religion	Relat	Grade	Ethnic	Marital	Cod EA
2349	amina bereo dule	3526	amina berou dule	10	2	4	4	99	1	1	2_1_5_8_88_1_8_9_1
2710	asiya hasen mehamed	2109	asiya hassan mohammed	30	2	4	2	99	1	2	2_1_5_8_88_1_8_9_1
2590	berao yaidu wagrıs	326	berou yaidu wagrıs	67	1	4	2	4	1	2	2_1_5_8_88_1_8_9_1
2831	fatuma barao halem	3307	fatuma bereou helem	19	2	4	4	99	1	1	2_1_5_8_88_1_8_9_1
2723	fatuma dawed mehamed	1892	fatuma dawud mohammed	18	2	4	2	99	1	2	2_1_5_8_88_1_8_9_1
3173	fatuma mehamed mukna	961	fatuma mohammed mukuna	23	2	4	2	99	1	2	2_1_5_8_88_1_8_9_1
2677	hamed fanta eissa	2344	hamed fanta ese	15	1	4	5	99	1	1	2_1_5_8_88_1_8_9_1
3400	hawa amin mehamed	444	hawa amin mohammed	16	2	4	4	99	1	1	2_1_5_8_88_1_8_9_1
3310	leko keloıyta gılsa	399	lokou koloıyta gılsa	23	2	4	2	99	1	2	2_1_5_8_88_1_8_9_1
3435	mesigo hasen ali	1261	mesigo hassen ali	16	2	4	1	99	1	2	2_1_5_8_88_1_8_9_1

Cliccando su “Linkage” (Figura 11), viene visualizzata la lista degli abbinamenti generati in modo automatico nella fase di elaborazione corrente. Nella schermata qui mostrata, è possibile passare alla fase successiva cliccando su “Step 3” oppure a quella precedente cliccando su “Step 1”.

Cliccando sul simbolo , il passo corrente di abbinamento automatico viene eseguito nuovamente.

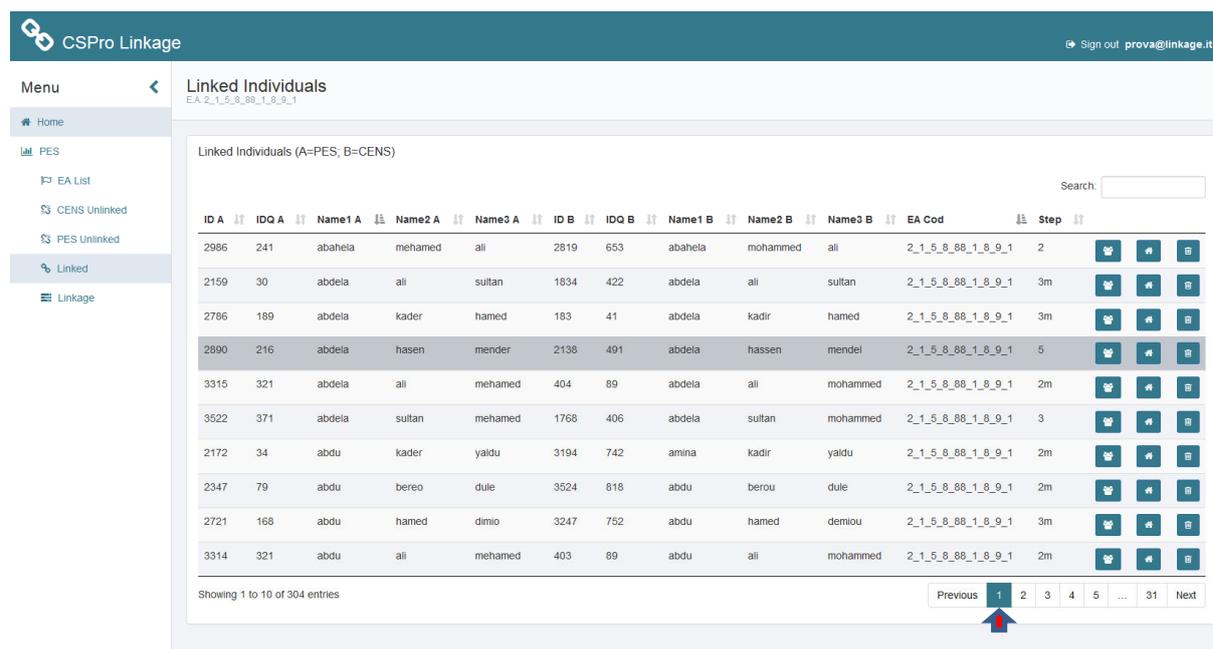
Figura 12. Ricerche su tabella

The screenshot shows the 'Linked Individuals' search results in the CSPro Linkage application. The title bar reads 'Linked Individuals' with the identifier 'EA_2_1_5_8_88_1_8_9_1'. The sidebar menu on the left has 'Linkage' highlighted. The main content area displays a table of linked individuals with a search bar at the top right containing the text 'hamedu'. The table has the following columns: ID A, IDQ A, Name 1 A, Name 2 A, Name 3 A, ID B, IDQ B, Name 1 B, Name 2 B, Name 3 B, EA Cod, and Step. The table contains 7 rows of data, with the 'Name 3 A' and 'Name 3 B' columns highlighted in red boxes. Below the table, it says 'Showing 1 to 7 of 7 entries (filtered from 304 total entries)' and includes 'Previous', '1', and 'Next' navigation buttons.

ID A	IDQ A	Name 1 A	Name 2 A	Name 3 A	ID B	IDQ B	Name 1 B	Name 2 B	Name 3 B	EA Cod	Step
2722	169	ali	sako	hamedu	1891	437	ali	seko	hamedu	2_1_5_8_88_1_8_9_1	2m
2817	197	awel	seko	mehamed	2644	611	awolo	seko	hamedu	2_1_5_8_88_1_8_9_1	1m
3576	382	aysha	hamedu	yaidu	3818	885	eysa	hamedu	yaidu	2_1_5_8_88_1_8_9_1	1m
3886	458	fatuma	hamedu	mehamed	2155	495	fatuma	hamedu	mohammed	2_1_5_8_88_1_8_9_1	4
3574	382	hamedu	yaidu	ali	3816	885	hamedu	yaidu	ali	2_1_5_8_88_1_8_9_1	1
2137	26	mehamed	alhamedu	hasen	1201	269	mohammed	alhamedu	hamed	2_1_5_8_88_1_8_9_1	4m
3577	382	zehara	hamedu	yaidu	3819	885	zehara	hamedu	yaidu	2_1_5_8_88_1_8_9_1	1m

È possibile applicare dei filtri alle righe di una tabella digitando in “Search” la stringa che si intende utilizzare per selezionare le righe (Figura 12). In questo modo saranno mostrate solo le righe che contengono la stringa in almeno una colonna. Tale funzione è disponibile per tutte le tabelle di dati dell’applicazione.

Figura 13. Paginazione di una tabella

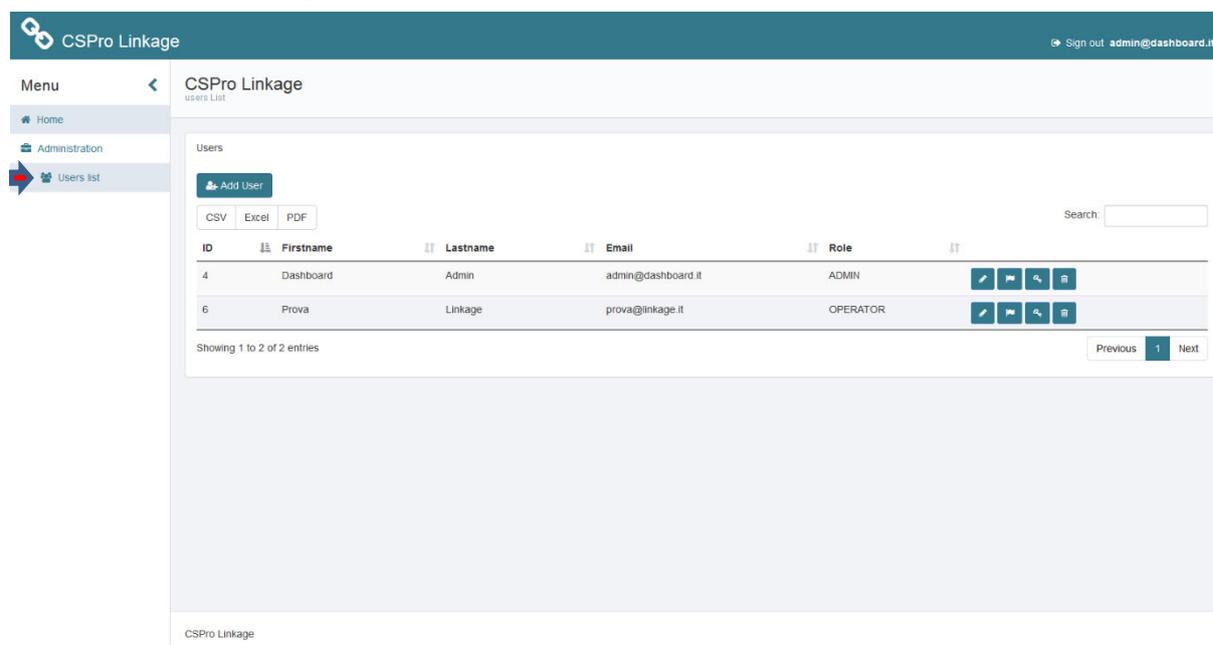


La paginazione di una tabella viene impiegata ogni qualvolta quest’ultima sia composta da più di dieci righe.

La pulsantiera in basso a destra permette di selezionare le pagine successive alla prima (Figura 13).

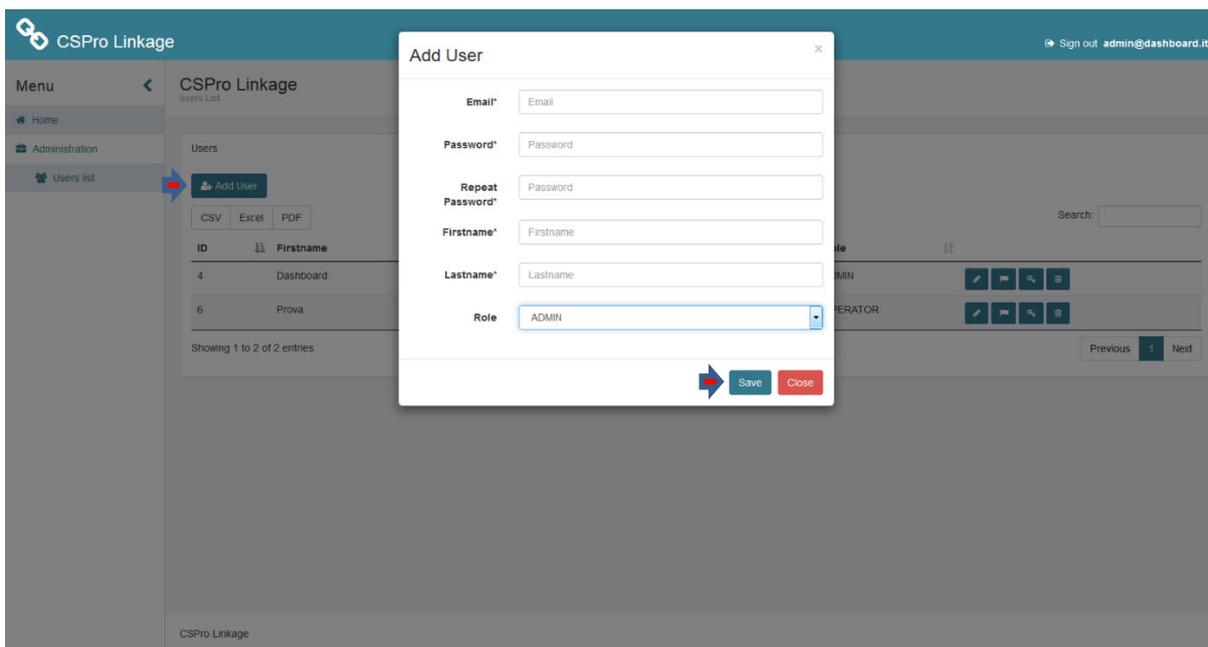
Interfaccia amministratore

Figura 14. Gestione degli utenti



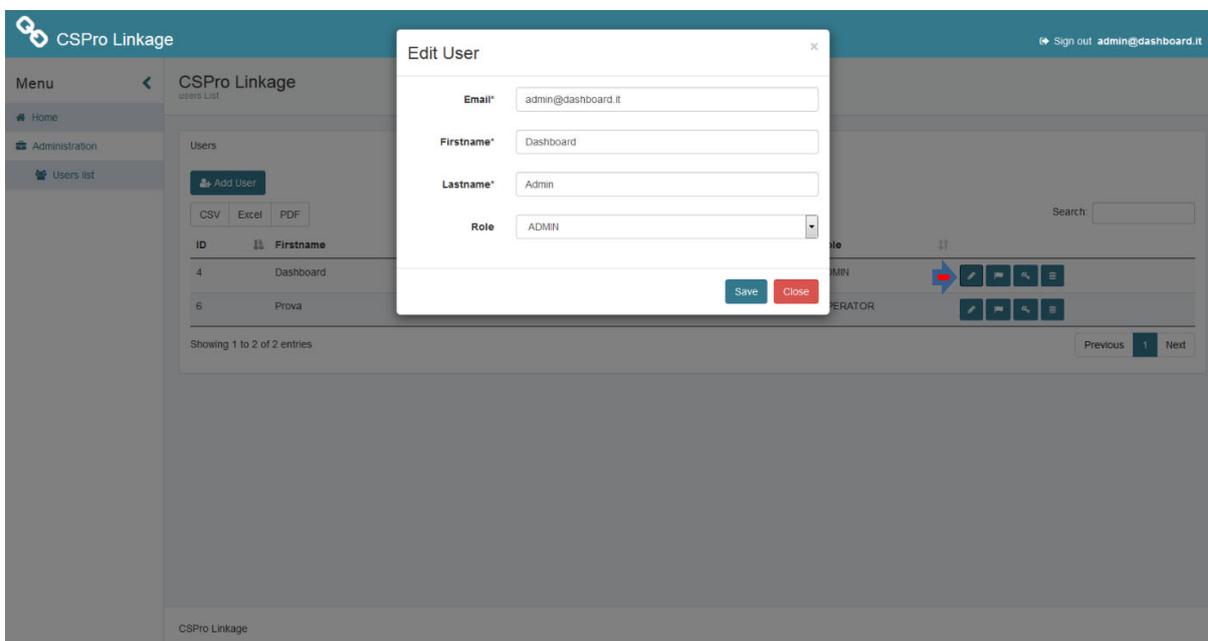
Cliccando su “User list” (Figura 14), è possibile gestire gli utenti.

Figura 15. Inserimento di un nuovo utente



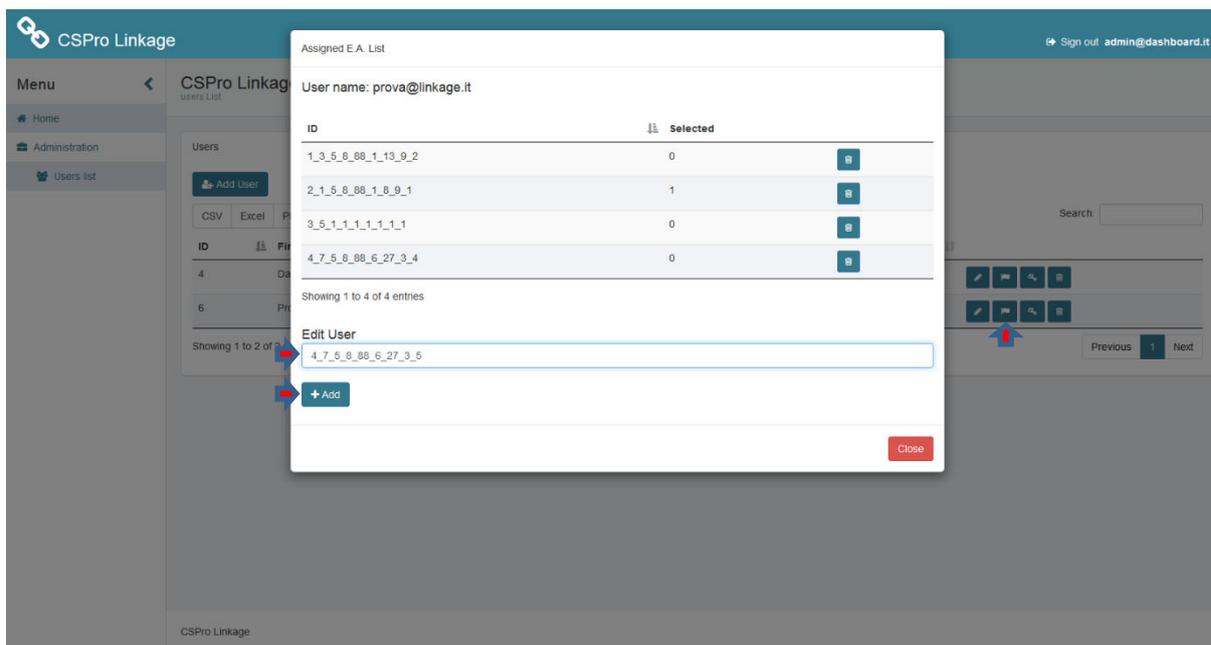
Per aggiungere un nuovo utente, cliccare su “Add User”, inserire i suoi dati e salvarli con il tasto “Save” (Figura 15).

Figura 16. Modifica del profilo di un utente già esistente



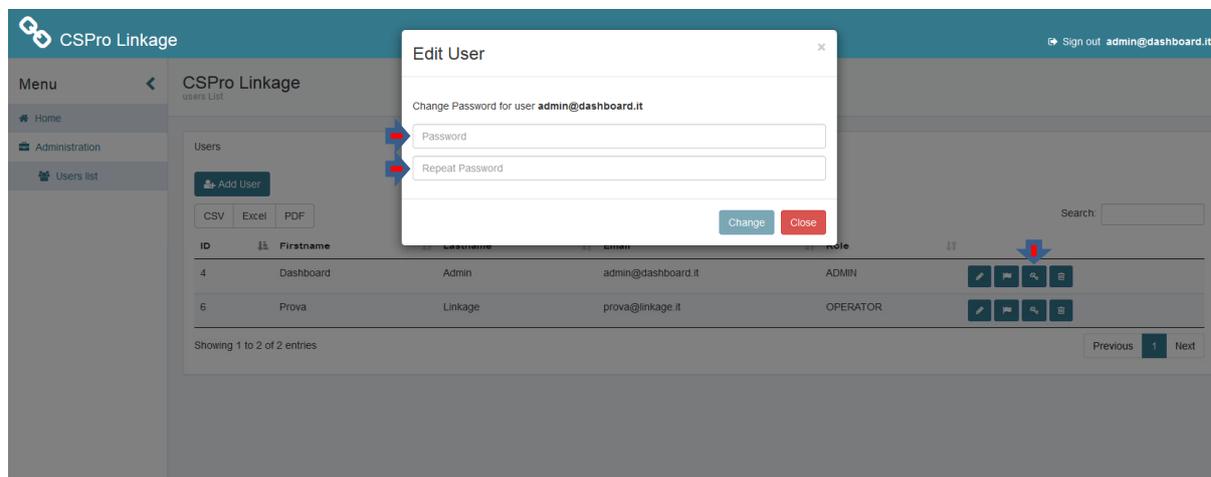
Cliccando sul simbolo della matita (Figura 16), è possibile modificare il profilo dell'utente della riga corrispondente.

Figura 17. Assegnazione di una o più sezioni di Censimento a un utente



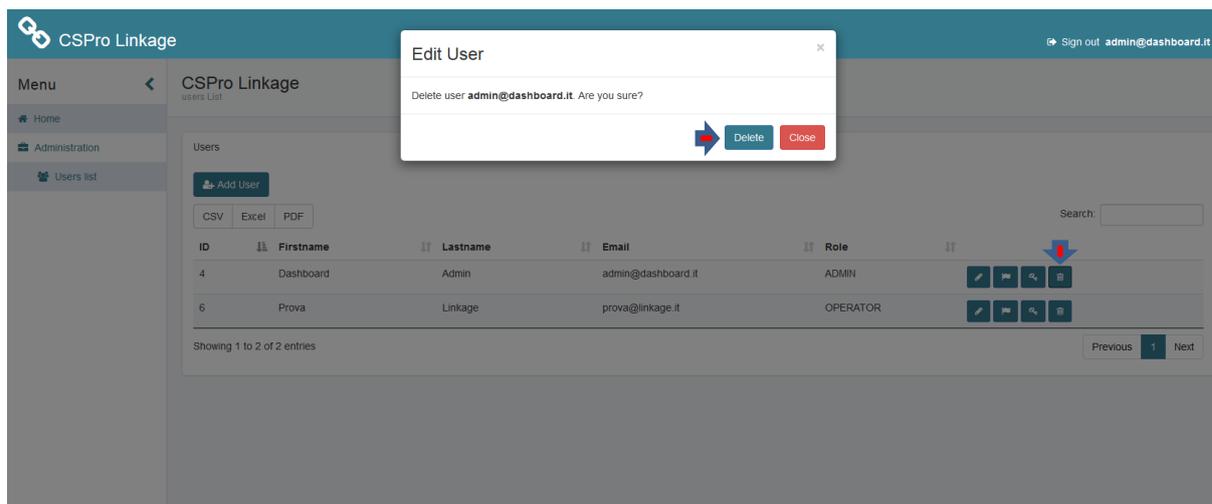
Cliccando sul simbolo della bandiera (Figura 17), è possibile gestire l’assegnazione delle sezioni di Censimento per l’utente corrispondente. Per assegnare una sezione a un utente, occorre inserire il codice identificativo della sezione in “**Edit User**” e cliccare su “**Add**”.

Figura 18. Modifica della password



Cliccando sul simbolo della chiave (Figura 18), è possibile modificare una password. All’apertura del *pop-up*, la nuova password potrà essere salvata con il tasto “**Change**”.

Figura 19. Cancellazione di un utente



Cliccando sul simbolo del cestino (Figura 19), si aprirà un *pop-up* dal quale con il tasto “Delete” è possibile cancellare un utente.

Riferimenti bibliografici

- Amarone, M., C. Colasanti, e T. Pellicciotti. 2016. “La valorizzazione della funzione informatica presso il Dipartimento di statistica giordano nell’attività di cooperazione italiana”. *Istat Working Papers*, N. 9/2016. https://www4.istat.it/it/files/2016/04/IWP_9_2016.pdf.
- Bruno, M., F. Grassia, J. Handley, A. Abayneh Abate, D. Deremew Mamo, and A. Girma. 2020. “Census metadata driven data collection monitoring: the Ethiopian experience”. *Statistical Journal of the IAOS*, Volume 36, Issue 1: 67–76.
- Cooperazione internazionale per lo sviluppo. 2017. *Documento triennale di programmazione e di indirizzo 2016 – 2018*. https://www.esteri.it/mae/resource/doc/2017/03/doc_triennale_2016-2018_-_finale_approvato.pdf.
- Ministero degli Affari Esteri e della Cooperazione internazionale - MAE, Direzione Generale per la Cooperazione allo Sviluppo. 2015. *La statistica come conoscenza indispensabile per la cooperazione, strategica per lo sviluppo. Piano d’Azione*. Roma: MAE.
- Statistics South Africa - STATSSA. 2012. “Post-enumeration survey (PES). Results and methodology”. *Report n. 03-01-46*. Pretoria, South Africa: STATSSA.
- United Nations, Department of Economic and Social Affairs, Statistics Division. 2017. *Principles and Recommendations for Population and Housing Censuses. Revision 3*. New York, NY, U.S.: United Nations.
- United Nations Secretariat, Department of Economic and Social Affairs, Statistics Division. 2010. *Post Enumeration Surveys: Operational Guidelines. Technical Report*. New York, NY, U.S.: United Nations.
- Wolter, K.M. 1986. “Some Coverage Error Models for Census Data”. *Journal of the American Statistical Association*, Volume 81, N. 394: 338-346.

Informazioni per le autrici e per gli autori

La collana è aperta alle autrici e agli autori dell'Istat e del Sistema statistico nazionale e ad altri studiosi che abbiano partecipato ad attività promosse dall'Istat, dal Sistan, da altri Enti di ricerca e dalle Università (convegni, seminari, gruppi di lavoro, etc.).

Coloro che desiderano pubblicare su questa collana devono sottoporre il proprio contributo al Comitato di redazione degli *Istat working papers*, inviandolo per posta elettronica all'indirizzo: iwp@istat.it.

Il saggio deve essere redatto seguendo gli standard editoriali previsti (disponibili sul sito dell'Istat), corredato di un sommario in Italiano e in Inglese e accompagnato da una dichiarazione di paternità dell'opera.

Per le autrici e gli autori dell'Istat, la sottomissione dei lavori deve essere accompagnata da un'e-mail della/del propria/o referente (Direttrice/e, Responsabile di Servizio, etc.), che ne assicura la presa visione.

Per le autrici e gli autori degli altri Enti del Sistan la trasmissione avviene attraverso la/il responsabile dell'Ufficio di statistica, che ne prende visione. Per tutte le altre autrici e gli altri autori, esterni all'Istat e al Sistan, non è necessaria alcuna presa visione.

Per la stesura del testo occorre seguire le indicazioni presenti nel foglio di stile, con le citazioni e i riferimenti bibliografici redatti secondo il protocollo internazionale 'Autore-Data' del *Chicago Manual of Style*.

Attraverso il Comitato di redazione, tutti i lavori saranno sottoposti a un processo di valutazione doppio e anonimo che determinerà la significatività del lavoro per il progresso dell'attività statistica istituzionale.

La pubblicazione sarà disponibile su formato digitale e sarà consultabile on line gratuitamente.

Gli articoli pubblicati impegnano esclusivamente le autrici e gli autori e le opinioni espresse non implicano alcuna responsabilità da parte dell'Istat.

Si autorizza la riproduzione a fini non commerciali e con citazione della fonte.