

istat working papers

N.6
2019

Censimento permanente: proposte per l'arricchimento dell'offerta informativa dell'output censuario

*Simone Ambroselli, Dario Ercolani, Silvana Garozzo, Evelina Paluzzi,
Roberta Rizzi, Alessandro Sasso, Mariangela Verrascina*

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	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	

Cura editoriale:

Vittorio Cioncoloni

Istat Working Papers

Censimento permanente: proposte per l'arricchimento dell'offerta informativa dell'output censuario

N. 6/2019

ISBN 978-88-458-1992-6

© 2019

Istituto nazionale di statistica
Via Cesare Balbo, 16 – Roma



Salvo diversa indicazione, tutti i contenuti pubblicati sono soggetti alla licenza Creative Commons - Attribuzione - versione 3.0.

<https://creativecommons.org/licenses/by/3.0/it/>

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

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

Censimento permanente: proposte per l'arricchimento dell'offerta informativa dell'output censuario

Simone Ambroselli, Dario Ercolani, Silvana Garozzo, Evelina Paluzzi,
Roberta Rizzi, Alessandro Sasso, Mariangela Verrascina¹

Sommario

Il Gruppo di lavoro interdipartimentale Archetipo (ARCHivi E sisTema di Indagini integrate per il censimento permanente della POPolazione) ha avuto il compito di definire il disegno strategico del Censimento della popolazione e delle abitazioni. Il Work Package D3 di Archetipo ha avuto il mandato di individuare, a partire da quelle già incluse in SIM (Sistema Integrato di Microdati), variabili di interesse statistico per l'arricchimento dell'offerta informativa che tradizionalmente caratterizza l'output censuario. A partire dalle indicazioni presenti nel mandato, le attività di ricerca hanno riguardato: la valutazione delle fonti amministrative sulla base dei metadati disponibili in SIM; la valutazione dei sistemi relazionali in fase di sviluppo nel SIM; la verifica della possibilità di rendere compatibili con gli obiettivi dell'indagine censuaria alcune delle proposte di output realizzate e in corso di realizzazione dal progetto Archimede (ARCHivio Integrato di Microdati Economici e DEMOSociali).

Parole chiave: censimento permanente; dati amministrativi; registri statistici; integrazione dati.

Abstract

Istat Working Group Archetipo (Archives and system of integrated surveys for the permanent census population) had the task of defining the strategic plan of the Census of Population and Housing. The Work Package D3 of Archetipo had the mandate to identify new statistical variables to enrich the traditional census output, starting from those already included in the Istat platform SIM (Integrated Micro data System). The activities of the WP, taking into account the basic instructions of the mandate, were focused on: the assessment of the administrative sources on the basis of the available SIM metadata; the evaluation of the relational systems under development in SIM; the assessment of the possibility to make consistent with the census targets some of the outputs realized (and under realization) by the Istat Project Archimede (Integrated archive of economic and demographic micro data).

Keywords: permanent census; administrative data; statistical registers; data integration.

¹ Il lavoro è stato svolto dai membri del Gruppo di Lavoro interno Archetipo: Simone Ambroselli, Lorenzo Cassata, Simone De Angelis, Dario Ercolani, Silvana Garozzo, Evelina Paluzzi, Roberta Rizzi, Alessandro Sasso, Mariangela Verrascina, Donatella Zindato.

Introduzione e conclusioni (paragrafo 5) sono state scritte da tutti gli autori. I paragrafi 2, 2.1 e 4 sono a cura di Simone Ambroselli; il paragrafo 2.2 è a cura di Dario Ercolani; il paragrafo 2.3 è a cura di Mariangela Verrascina; i paragrafi 3 e 3.1 sono a cura di Alessandro Sasso; i paragrafi 3.2 e 3.2.1 sono a cura di Silvana Garozzo; i paragrafi 3.3, 3.3.1 e 3.3.2 sono a cura di Roberta Rizzi e Evelina Paluzzi.

Gli articoli pubblicati impegnano esclusivamente gli Autori, le opinioni espresse non implicano alcuna responsabilità da parte dell'Istat.

Indice

	Pag.
1. Introduzione	5
2. Decisioni operative	6
2.1 Linee di ricerca – SIM	8
2.2 Linee di ricerca – Analisi dei risultati del progetto Archimede.....	10
2.3 Linee di ricerca – Analisi della diffusione dei dati censuari in alcuni Stati esteri.....	11
3. Sezioni e possibili sviluppi	14
3.1 Informazioni anagrafiche e abitazioni	14
3.2 Edifici	16
3.2.1 Simulazione di analisi comparativa tra dati catastali e dati censuari.....	17
3.3 Popolazione straniera	18
3.3.1 Criticità del censimento della popolazione straniera.....	18
3.3.2 Prime proposte	19
4. Ulteriori proposte di approfondimento	21
5. Criticità e prospettive	23

1. Introduzione

La realizzazione dei censimenti permanenti, caratterizzati dall'utilizzo integrato di fonti amministrative e dallo svolgimento di rilevazioni statistiche a supporto, necessita di un'adeguata fase di progettazione utile per definire concretamente le strategie da adottare per assolvere gli obblighi di legge e migliorare, contestualmente, l'offerta informativa dell'Istat. Al fine di definire il disegno strategico del Censimento della popolazione e delle abitazioni, è stato costituito, all'interno dell'Istituto, il Gruppo di lavoro interdipartimentale Archetipo (ARCHivi E sisTema di Indagini integrate per il censimento permanente della POpolazione). Il gruppo di lavoro ha operato per circa un anno a partire dalla metà del 2015. Operativamente le attività di ricerca sono state svolte da sei sottogruppi: tre orientati al macro-obiettivo "conteggio della popolazione"; altri tre specifici per la stima delle variabili socio-economiche. Riguardo questo secondo obiettivo, ai tre *Work Package* (WP) sono stati affidati compiti diversi, nell'ordine:

- a) classificazione delle variabili, ridefinizione del questionario e dell'output censuario;
- b) individuazione delle variabili già sostituibili con dati di fonte amministrativa e/o Registri statistici e definizione dei modelli di stima;
- c) arricchimento dell'offerta informativa tradizionale del censimento.

Il presente documento sintetizza i risultati del sottogruppo WPD3, il cui obiettivo è stato quello di individuare variabili di interesse statistico per l'arricchimento dell'offerta informativa che tradizionalmente caratterizza l'output censuario. Le condizioni da considerare per l'ampliamento dell'offerta informativa erano le seguenti:

- individuare "informazioni" con un livello di dettaglio caratteristico di un'indagine censuaria;
- descrivere le singole unità con caratteristiche, ad oggi, non presenti nei censimenti.

Punto di partenza per le attività di ricerca è stato il SIM (Sistema Integrato di Microdati), una struttura informativa di base su individui, famiglie e unità economiche realizzata mediante l'integrazione concettuale e fisica dei microdati acquisiti da fonti amministrative e statistiche (Runci, Di Bella, Galiè, 2016; Ambroselli, 2015). Le fonti statistiche disponibili durante il periodo di lavoro erano costituite dai microdati definitivi dell'ultima tornata censuaria del 2011 e dai flussi informativi periodici su trasferimenti di residenza, iscritti all'anagrafe per nascita e cancellazioni anagrafiche per morte, utilizzati correntemente dall'Istituto per la produzione delle statistiche demografiche².

Le decisioni operative e metodologiche adottate dal WP per la formalizzazione delle proposte di contenuti informativi nuovi sono descritte nel §2.

Per la realizzazione del presente report sono state analizzate sia le singole fonti amministrative sia i sistemi relazionali sviluppati nel SIM (§ 2.1). Ulteriori linee di ricerca hanno riguardato la valutazione dei risultati del progetto Archimede³ (Garofalo, 2014; § 2.2) e l'analisi della diffusione dei dati censuari in alcuni Stati esteri (§ 2.3).

Le aree tematiche del censimento analizzate nel dettaglio sono state le seguenti: alloggi e famiglie (§ 3.1), edifici (§ 3.2) e un focus specifico sulla popolazione straniera (§ 3.3). Altre aree tematiche, per le quali era già possibile, almeno parzialmente, sviluppare dei modelli di stima in termini di sostituibilità delle variabili, sono state esaminate nell'ambito di altri sottogruppi di Archetipo (istruzione e formazione; condizione professionale e lavoro; pendolarismo/mobilità sul territorio).

La formalizzazione di ulteriori proposte di approfondimento è inserita nel §4, mentre, nel successivo paragrafo (§5) sono evidenziate le principali criticità e prospettive.

² Portale delle statistiche demografiche <http://demografiche.istat.it/index.php?id=11>.

³ ARCHivio Integrato di Microdati Economici e DEMografici: progetto Istat che ha svolto delle sperimentazioni nel periodo 2013-2016 al fine di ampliare l'offerta informativa dell'Istat mediante la produzione di collezioni di dati elementari.

2. Decisioni operative

Nell'ambito del sottogruppo, l'arricchimento dell'offerta informativa è stato inteso non come sola individuazione di singole "variabili di interesse statistico". In realtà, contenuti informativi nuovi, sulla base delle fonti e dei sistemi disponibili, possono essere intesi anche come relazioni tra unità e/o combinazioni di più variabili. Inoltre, una delle principali novità che presenta il censimento permanente è la produzione annuale di dati e questo permette di valutare la significatività di alcune informazioni da un punto di vista longitudinale.

A partire dalle indicazioni presenti nel mandato, le attività di ricerca potenzialmente utili, individuate all'interno del WP D3, possono essere ricondotte ai seguenti filoni principali:

1. la valutazione delle fonti disponibili in SIM sulla base dei metadati (variabili e modalità) attualmente presenti nel sistema integrato; inoltre, sono state valutate, sulla base delle informazioni disponibili, anche delle fonti amministrative specifiche non ancora nel sistema ma di prossima acquisizione;
2. la valutazione dei sistemi relazionali sviluppati (e in fase di sviluppo) nel SIM;
3. la verifica della possibilità di rendere compatibili alcune delle proposte di output realizzate (e in corso di realizzazione) dal progetto Archimede con gli obiettivi dell'indagine censuaria;
4. l'analisi del piano di diffusione nazionale e di altre esperienze internazionali per valutare la possibilità di proporre eventuali contenuti informativi "nuovi" espressi a livello locale, nazionale o dalla comunità scientifica.

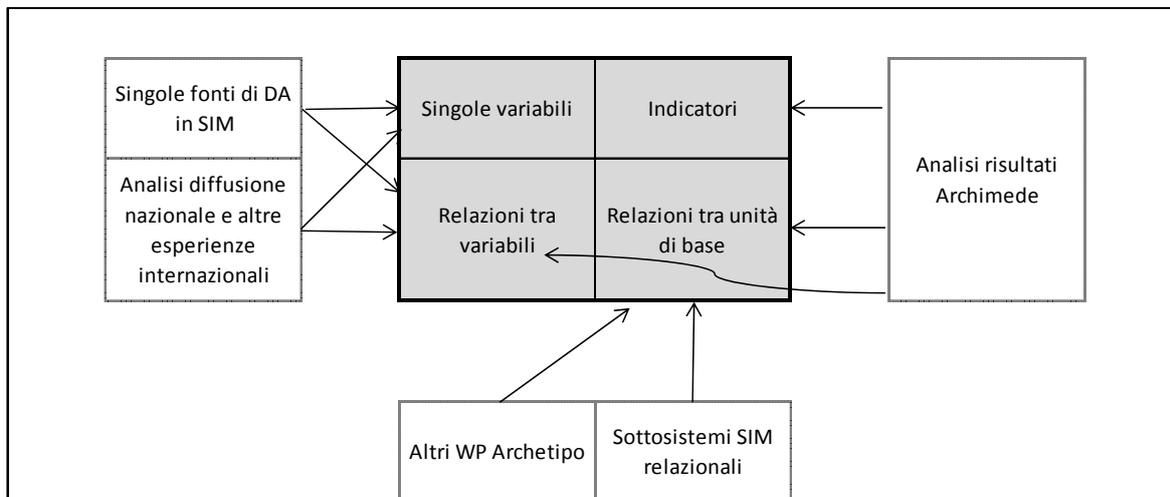
Sulla base delle ipotesi di lavoro iniziali, a queste linee di ricerca, si sarebbe potuta aggiungere l'analisi di eventuali *feedback* da parte di altri WP del gruppo di lavoro, soprattutto riguardo l'identificazione di popolazioni e sottopopolazioni specifiche. Da sottolineare come all'interno del Gruppo di lavoro Archetipo, un altro WP abbia svolto analisi specifiche riguardo i segnali di dimora abituale in Italia e l'individuazione di sottopopolazioni critiche⁴.

In Figura 1 è mostrato il collegamento tra le linee di analisi e le diverse tipologie di contenuti informativi "nuovi" potenzialmente individuabili. Come detto in precedenza, oltre a singole variabili da includere nell'output censuario, la ricerca di contenuti "nuovi" ha riguardato le relazioni tra variabili, anche sotto forma di indicatori, e quelle tra unità. Da sottolineare come, per il presente lavoro, con Dataset Amministrativo (DA) si intenda un insieme strutturato di dati estratti da una o più fonti amministrative, prima di qualsiasi trattamento o validazione statistica (in ambito Sistan)⁵.

⁴ Borrelli F., Chieppa A., Di Domenico S., Gallo G., Rosati S., Tomeo V. (2016). Primi risultati della sperimentazione condotta su fonti amministrative capaci di valutare i segnali di dimora abituale in Italia e l'individuazione di sottopopolazioni critiche F. Roma: Istat Working Papers 23 2016.
https://www4.istat.it/it/files/2018/02/IWP_23_2016.pdf?title=Istat+Working+Papers+n.+23%2F2016+-+28%2Ffeb%2F2018+-+IWP_23_2016.pdf.

⁵ Glossario Istat. Adattamento del Glossario prodotto nell'ambito del progetto internazionale ESSnet AdminData, "Glossary of the main terms related to the use of admin data in producing business statistical", 2012, https://ec.europa.eu/eurostat/cros/content/glossary-main-terms-related-use-admin-data-producing-business-statistics-first-release_en.

Figura 1 – Linee di ricerca e tipologie di potenziali contenuti informativi “nuovi”



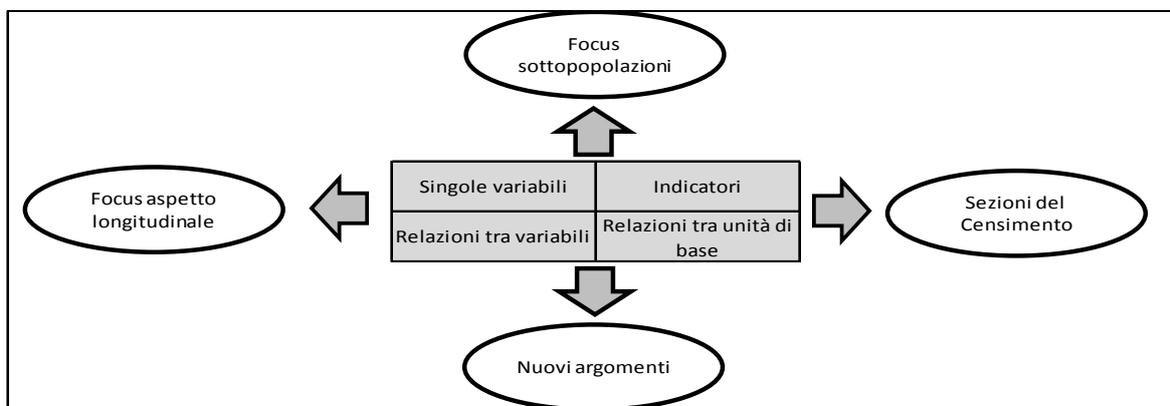
Fonte: Elaborazione degli autori per la fase di individuazione di fonti e sistemi disponibili

Rispetto al SIM, inteso sia come singole fonti gestite sia come sistemi relazionali, la prima fase di analisi è stata rivolta all’individuazione di singole variabili, relazioni tra variabili e relazioni tra unità di base che possano essere considerate come dei contenuti nuovi da diffondere nel Censimento. L’ipotesi emersa all’interno del WP è che, accanto al SIM, alcune indicazioni sulle variabili e, soprattutto, sulle relazioni tra variabili, possano essere ricavate dal confronto della diffusione italiana con quella che altri Paesi hanno messo a disposizione degli utenti a livello nazionale per il Censimento del 2011. Il progetto Archimede è, invece, potenzialmente utile per identificare degli indicatori efficaci e sintetici per descrivere fenomeni socio-economici con le caratteristiche del dettaglio censuario. Inoltre, per le specifiche ricerche svolte da alcuni dei progetti Archimede, possono essere tratte delle indicazioni utili in termini di relazioni tra unità e variabili.

In questa fase, ogni linea di ricerca è stata vista come una fonte di input da cui trarre idee, spunti, suggestioni da industrializzare per produrre contenuti “nuovi”, anche provando a rendere compatibile con il Censimento informazioni colte in altri ambiti.

Il passo successivo, quasi contestuale alla prima fase di analisi, è stato quello di incanalare gli input sulle potenziali informazioni “nuove” in ambiti specifici (Figura 2).

Figura 2 – Dai singoli contenuti ai focus specifici



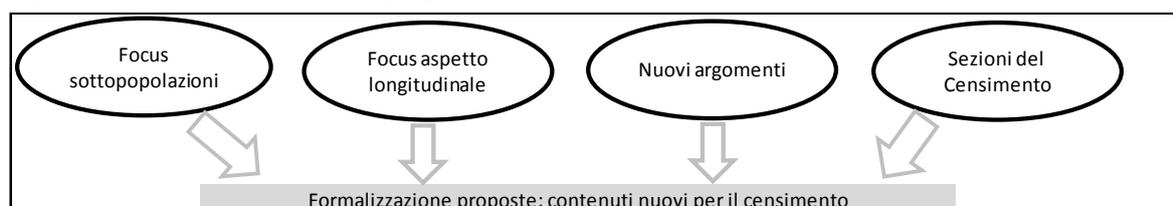
Fonte: Elaborazione degli autori per la fase di identificazione dei focus informativi

Il potenziale informativo “nuovo” trovato nella fase precedente, in termini di variabili, relazioni (tra variabili o unità) e indicatori, è stato attribuito, ove possibile, alle sezioni esistenti del censi-

mento. In questo modo si è potuto distinguere tra argomenti del tutto nuovi, eventualmente da proporre, e quelli già presenti nell'output censuario che, alla luce dei risultati della ricerca, potevano essere, almeno parzialmente, ricavati dall'integrazione di dati amministrativi. In effetti, come sarà esposto successivamente, sono emerse delle potenzialità informative per alcune sottopopolazioni e per gli aspetti longitudinali.

Gli ambiti specifici sono stati successivamente rivalutati in un'ottica complessiva al fine di formalizzare le proposte rendendole omogenee e evitando duplicazioni di attività (Figura 3).

Figura 3 – Formalizzazione delle proposte



Fonte: Elaborazione degli autori per la fase di formalizzazione delle proposte

In quest'ultima fase si è provato a scendere sempre più nel dettaglio informativo passando dagli argomenti alle sottopopolazioni coinvolte e alle singole variabili nuove che possono essere proposte.

2.1 Linee di ricerca – SIM

Il SIM è il *repository* dei dati amministrativi acquisiti dall'Istat. Ha lo scopo di supportare trasversalmente i processi di produzione dell'istituto e “oltre alle unità di base, che allo stato attuale comprendono le tre tipologie di Individui, Unità economiche e Luoghi, anche le relazioni tra le unità dello stesso tipo o di diverso tipo costituiscono uno specifico interesse per l'analisi dei fenomeni statistici” (Runci, Di Bella, Galiè, 2016).

Gli obiettivi del processo di integrazione del SIM sono:

- identificare ogni oggetto (individui; unità economiche; loro relazioni) in fonti diverse con un codice univoco e stabile nel tempo;
- definire, per ogni oggetto, le relazioni logiche e fisiche, nel tempo e nello spazio, tra le informazioni disponibili da fonti diverse.

Il risultato finale è costituito dalla realizzazione di strutture di dati con unità elementari appartenenti a popolazioni statistiche utili per la realizzazione di Registri e sottosistemi informativi (Ambroselli, 2015). “SIM individui” e “SIM unità economiche” sono due dei sottosistemi di base e per gli oggetti presenti nei due sottosistemi sono assegnate, rispettivamente, le due chiavi primarie, il “codice individuo” e il “codice unità”. Dopo la fase di integrazione, tutte le fonti incluse nel sistema conterranno i codici identificativi univoci del SIM. Le fonti presenti nel SIM con dati/segnali riguardanti gli individui esaminate nell'ambito del WP sono quelle indicate nella Tabella 1.

Tabella 1 – SIM Individui – fonti di dati amministrativi incluse (fine 2015)

COMUNI - Liste anagrafiche comunali	INPS - Archivio delle denunce retributive mensili - UniEmens
AGENTRATE - Anagrafe delle Persone Fisiche	INPS - Beneficiari CIG a pagamento diretto
ACI - Archivio ACI	INPS - Beneficiari e titolari Social Card
AGENTRATE - Dichiarazioni 730	INPS - Casellario dei Pensionati
AGENTRATE - Modello Unico	INPS - DMAG unico
CCIAA - Archivio Persone d'Impresa	INPS - Lavoratori autonomi in agricoltura
CCIAA - Archivio soci delle società di capitale (Banca Dati Soci)	INPS - Posizioni assicurative (INPDAP)
CONSODATA - Archivio dei contratti di telefonia fissa - PRIVATI	INPS - Rapporti di lavoro domestico
ENPALS - Imprese e iscritti contribuenti sport professionistico e spettacolo	INPS - Trattamenti non pensionistici
ENPAM - Iscritti all'albo dei medici chirurghi e degli odontoiatri	MAE - AIE Archivio consolare degli italiani all'estero
INAIL - DNA	MEF - Anagrafica del personale per tutti gli enti
INAIL - Lavoratori interinali	MEF - Banca dati reddituale
INAIL - Lavoratori parasubordinati	MEF - Modello 770 statistico
INPGI - Dati sulle imprese e sugli iscritti - Collaboratori	MINISTERO DELL'INTERNO - Permessi di soggiorno
INPGI - Dati sulle imprese e sugli iscritti - Gestione principale	MIUR - Anagrafe degli studenti
INPGI - Dati sulle imprese e sugli iscritti - Liberi professionisti	MIUR - Anagrafe degli studenti - Esiti scolastici
INPS - Archivio dei lavoratori autonomi: artigiani e commercianti	MIUR - Archivio degli iscritti e delle iscrizioni universitarie
INPS - Archivio dei parasubordinati collaboratori	MIUR - Archivio dei laureati
INPS - Archivio dei parasubordinati professionisti	MIUR - Archivio del personale delle scuole statali
	MIUR - Dati sul personale docente e non docente universitario

Fonte: Elaborazione degli autori sulla base dei metadati SIM

Da sottolineare come il SIM Individui includa anche le fonti statistiche censuarie (2011) e i flussi informativi demografici: Iscritti all'Anagrafe per nascita; Cancellazioni anagrafiche per morte; Trasferimenti di residenza.

E' stata valutata anche una fonte di dati amministrativi, al momento, non presente nel SIM: il Catasto degli edifici (ente fornitore: Agenzia delle Entrate) che si è rivelata particolarmente utile per le analisi relative ai temi "abitazioni" (vedi successivo §3.1) e "edifici" (§3.2). Infine, sono stati analizzati alcuni specifici DA utilizzati nel sottosistema SIM delle unità economiche al fine di identificare contenuti informativi utili per gli individui e, eventualmente, le famiglie. In particolare, sono stati esaminati i seguenti DA.

- a. Agenzia delle Entrate – Anagrafe delle Persone Giuridiche: archivio contenente tutte le informazioni dei possessori di Partita Iva.
- b. CCIAA – Registro Imprese: nel Registro Imprese sono registrati i dati di tutte le imprese, e degli altri soggetti previsti dalla legge, con sede o unità locali sul territorio nazionale. Contiene informazioni anche su amministratori e soci delle imprese registrate.
- c. Consob – Archivio delle società quotate in borsa: archivio specifico per le società quotate con informazioni sui legami societari;
- d. Agea – Fascicolo aziendale: il fascicolo aziendale è unico per ciascun soggetto registrato (operatore agricolo) nell'Anagrafe del Settore Primario (Sistema Informativo Agricolo Nazionale - SIAN).
- e. Inps – Lavoratori Extracomunitari: archivio anagrafico specifico per lavoratori extracomunitari.

Nel SIM sono stati sviluppati anche dei sistemi relazionali tra le unità di base utili per valutare possibili strategie specifiche di diffusione su sottopopolazioni e/o unità complesse. Infatti, combinando le tre unità di base del sistema, Individui, Unità economiche e Luoghi, è possibile ottenere i sottosistemi delle relazioni. Al momento, è possibile costruire i sistemi relazionali elencati qui di seguito (Ambroselli, 2015).

- SIM relazioni tra unità economiche: obiettivo del sottosistema è quello di cogliere alcune possibili relazioni tra unità economiche.
- SIM relazioni tra individui: lo scopo principale è identificare le relazioni tra gli individui.
- SIM relazioni tra individui e unità economiche: integra le informazioni sulle relazioni tra gli individui e le unità economiche.
- SIM luoghi unità economiche: contiene le localizzazioni delle unità economiche presenti

nelle fonti amministrative.

- SIM luoghi individui: sono presenti gli indirizzi collegabili alle persone fisiche riconosciute dai DA inseriti nel sistema, quali, ad esempio, residenza anagrafica, domicilio fiscale, indirizzi delle utenze domestiche e così via.

Ai fini del presente lavoro, risultano di particolare interesse le relazioni individui-unità economiche, quelle tra individui e il sottosistema luoghi-individui. Nel primo caso, il relativo sottosistema contiene il collegamento tra individui e unità economiche ed è possibile cogliere le seguenti macrotipologie relazionali: lavoro; studio; ruolo societario. Da sottolineare come la presenza dei codici identificativi univoci del SIM in alcuni dei principali Registri statistici dell'Istat (ad esempio, Registro statistico delle imprese attive - Asia, Registro statistico Asia-Occupazione⁶, Anvis - ANagrafe Virtuale Statistica) permetta di collegare variabili e relazioni non solo alle fonti amministrative di base ma anche ad altri registri consolidati nell'Istituto, ampliando le possibilità informative.

Tramite il sottosistema “relazioni tra individui” è, invece, possibile collegare i diversi concetti di famiglia anagrafica e fiscale per ottenere l'insieme più ampio possibile di relazioni.

Infine, lo sviluppo del sottosistema sui luoghi degli individui può consentire di ampliare le possibilità di esaminare il fenomeno della mobilità sul territorio.

2.2 Linee di ricerca – Analisi dei risultati del progetto Archimede

Il Progetto Archimede (Garofalo, 2014) ha avuto l'obiettivo di ampliare l'offerta informativa dell'Istat mediante la produzione di collezioni di dati elementari di tipo longitudinale (ad es. percorsi sociali ed economici, carriere lavorative) e *cross section* da rendere disponibili all'utenza, utili alla ricerca sociale ed economica, alla programmazione territoriale e settoriale, alla valutazione delle politiche pubbliche a livello nazionale, regionale e locale. Tale obiettivo è stato perseguito attraverso lo sfruttamento dei contenuti informativi delle basi di dati amministrative contenute nel SIM. Le prime sperimentazioni progettate e condotte hanno riguardato i seguenti temi:

- “Popolazione insistente / *Persons&Places*”⁷: identificazione, classificazione e quantificazione della popolazione che usa il territorio per motivi di lavoro e studio;
- “Precarietà lavorativa”⁸: identificazione, classificazione e qualificazione dei lavoratori con contratti di lavoro definiti precari;
- “Condizioni socio-economiche delle famiglie”⁹: analisi della condizione socio-economica delle famiglie residenti al fine di descriverne il grado di vulnerabilità;
- “Percorsi di istruzione, formazione e inserimento lavorativo”¹⁰: analisi dei percorsi formativi e di istruzione, dei fenomeni legati all'abbandono formativo e gli inserimenti occupazionali di soggetti aventi età compresa tra i 14 e i 29 anni.

Il fine delle sperimentazioni è stato quello, da un lato, di valutare l'effettiva potenzialità del progetto, dall'altro, di proporre e valutare la fattibilità di specifici prodotti statistici e dei sistemi di diffusione di output informativi. A tal fine è stata realizzata una base dati integrata contenente informazioni che ha permesso di ottenere, per ogni tema, i seguenti risultati:

- “Popolazione insistente”: costruzione della matrice origini/destinazioni della mobilità quotidiana per motivi di lavoro/studio per ogni comune utilizzando esclusivamente microdati ammini-

⁶ Struttura e occupazione delle imprese. Registro statistico delle imprese attive (Asia) <http://www.istat.it/it/archivio/188233> (periodo di riferimento anno 2014)

⁷ Principali fonti integrate: Liste Anagrafiche Comunali (LAC); Anagrafe Tributaria; archivi MIUR; dati sull'occupazione (Sistema Informativo Archimede Precarietà Lavorativa).

⁸ Principali fonti integrate: ASIA Occupazione; Posizioni Assicurative Ex Inpdap; Archivio Lavoratori Autonomi Agricoltura; Archivio Lavoro Domestico; Agenzia Entrate Modello Unico e Modello 770.

⁹ Principali fonti integrate: Censimento della Popolazione; Liste Anagrafiche Comunali (LAC); Banca Dati Redditi MEF; Agenzia Entrate Modello Unico; Casellario Centrale dei Pensionati; Trattamenti Monetari Non Pensionistici; archivi MIUR; dati sull'occupazione (Sistema Informativo Archimede Precarietà Lavorativa).

¹⁰ Principali fonti integrate: Censimento della Popolazione; popolazione insistente (vedi descrizione sottoprogetto *Persons&Places*); archivi MIUR; dati regionali su Istruzione e Formazione Professionale, tirocini e apprendistato (Comunicazioni Obbligatorie); dati sull'occupazione (Sistema Informativo Archimede Precarietà Lavorativa).

strativi; quantificazione all'interno della popolazione iscritta in Anagrafe dei residenti statici/dinamici e dei pendolari giornalieri/periodici; quantificazione sul territorio di ciascun comune della popolazione insistente, della popolazione dinamica (in ingresso/uscita), della popolazione statica; costruzione di indicatori di utilizzo del territorio quali: indice di attrazione di un territorio, rapporto di coesistenza tra individui insistenti e residenti.

- “Precarietà lavorativa”: realizzazione di un sistema informativo costituito dalla totalità degli occupati presenti negli archivi amministrativi – prevalentemente previdenziali e fiscali – classificati in lavoratori stabili e precari¹¹. L'esame del suddetto sistema permette analisi sia *cross section* sia longitudinali, nonché la costruzione di indicatori tra i quali: tasso di precarietà tra i nuovi occupati, tasso di permanenza nella condizione di precarietà, tasso di stabilizzazione.

- “Condizioni socio-economiche delle famiglie”: costruzione di un sistema informativo contenente le famiglie residenti e le loro caratteristiche strutturali, di istruzione, reddituali, occupazionali, al fine di: poterle classificare, effettuare analisi reddituali, costruire indicatori (ad es. percentuale di famiglie con bassa intensità lavorativa, percentuale di famiglie con giovani adulti, percentuale di famiglie monoreddito con minori), costruire indici sintetici di vulnerabilità sociale e individuare legami fiscali tra nuclei familiari.

- “Percorsi di istruzione, formazione e inserimento lavorativo”: implementazione di un dataset costituito da individui di età 14-29 anni presenti sul territorio – mediante segnali di residenza/studio/lavoro – attraverso il quale effettuare analisi trasversali e longitudinali sui percorsi di istruzione, abbandono scolastico e inserimento lavorativo.

Ai fini dei lavori del WP, l'esperienza del progetto Archimede ha rappresentato certamente un esempio rilevantissimo e concreto di utilizzo di dati fonte diversa, amministrativa e statistica, da cui trarre spunti metodologici da sfruttare nell'ottica censuaria. Le sperimentazioni in atto sono sicuramente inquadrabili tra le esperienze di produzione di output statistico di tipo *register-based* (Wallgren, A., Wallgren, B., 2007) che sfruttano appieno le potenzialità della micro-integrazione (“*micro-integration is the method that aims at improving the data quality in combined sources by searching and correcting for the errors on unit level*”, Bakker B. F.M., 2011). Allo stesso modo, i possibili sviluppi presentati nel successivo paragrafo 3 di questo lavoro, avendo come punti di partenza comuni il SIM e i DA amministrativi ad esso collegati, consentirebbero di ottenere risultati a livello di micro-dati.

2.3 Linee di ricerca – Analisi della diffusione dei dati censuari in alcuni Stati esteri

I Regolamenti europei emanati tra il 2008 e il 2010 per la tornata censuaria del 2010/2011 hanno imposto l'obbligo di fornire i risultati a Eurostat (attraverso il *Census Hub*¹²), con scadenza pre-stabilita. Allo stesso tempo, l'Italia ha dovuto soddisfare il fabbisogno informativo degli utenti nazionali e garantire continuità alle serie storiche censuarie. Di conseguenza, per il censimento del 2011 Istat ha predisposto due differenti Piani di diffusione, uno definito sulla base del Regolamento europeo¹³, l'altro in linea con quanto prodotto in Italia per i censimenti precedenti.

Istat ha avviato una ricognizione di quanto è stato effettuato negli altri Paesi a livello nazionale, in aggiunta a quanto previsto dal regolamento comunitario, con l'obiettivo di verificare se gli altri contesti europei avessero avuto necessità di “sdoppiare” la diffusione, così come accaduto in Italia.

¹¹ Sebbene non esista a livello internazionale una definizione univoca e formalizzata di precarietà lavorativa, soprattutto per motivi connessi alla disomogeneità tra normative nazionali in tema di mercato del lavoro, nell'ambito del Progetto si definisce lavoratore “precario” colui che in un determinato periodo di tempo, svolge esclusivamente attività di lavoro dipendente a termine, rapporti di collaborazione e lavoro autonomo connesso ad attività aventi caratteristiche di subordinazione.

¹² Il *Census Hub* è lo strumento creato da Eurostat per fornire agli utenti un facile accesso ai dati censuari dettagliati, confrontabili tra gli Stati membri, strutturati in maniera omogenea e armonizzati nei contenuti. (<https://ec.europa.eu/CensusHub2>)

¹³ COMMISSION REGULATION (EU) No 519/2010 of 16 June 2010 adopting the programme of the statistical data and of the metadata for population and housing censuses provided for by Regulation (EC) No 763/2008 of the European Parliament and of the Council.

Pertanto sono stati presi in considerazione alcuni Paesi esteri,¹⁴ con differenti approcci censuari, per esplorare la diffusione nazionale relativa al censimento 2011, in base alla tecnica adottata, e confrontarla con la produzione censuaria italiana. Lo scopo della ricognizione è stato di individuare nuove forme e strategie di diffusione da adottare nei censimenti futuri, oltre che scoprire possibili nuovi contenuti informativi¹⁵ per il censimento permanente.

I Paesi considerati nell'analisi sono stati classificati per tipo di censimento adottato nel 2011 e suddivisi in: censimento basato su registri, censimento combinato e censimento tradizionale. Nel primo caso si tratta di censimento effettuato utilizzando esclusivamente informazioni contenute nei registri e nelle altre fonti amministrative, nel secondo caso si tratta di un approccio che combina informazioni da registri e informazioni che provengono da indagini (esaustive o campionarie), infine, nel terzo caso si tratta del metodo tradizionale di enumerazione totale della popolazione.

I Paesi oggetto della ricognizione sono stati:

- Paesi Nordici (Danimarca, Norvegia, Finlandia, Svezia), Austria, Slovenia per il *Register-based census* (UNECE, 2007);
- Paesi Bassi, Spagna, Germania, Israele¹⁶ per il *Combined census*;
- Inghilterra e Galles, Italia per il *Traditional census*.

L'analisi della produzione censuaria degli altri Paesi è stata effettuata considerando la presenza di diffusione a livello nazionale di output censuario, il tipo di diffusione nazionale, i nuovi contenuti (rispetto a quanto diffuso in Italia) e le modalità di diffusione adottate a livello nazionale.

Alcune informazioni sui Paesi presi in considerazione sono state sintetizzate nello Schema 1: presenza di questionario di censimento, popolazione totale risultante al censimento, data di censimento, presenza di diffusione a livello nazionale e tipo di censimento.

Schema 1 – Paesi oggetto della ricognizione e tipologia di censimento

Area Geografica	Questionario	Totale Popolazione	Census Day	Diffusione nazionale	Tipo di censimento
Denmark	no FORM	5.560.628	31/12/2010	NO	REGISTER-BASED
Norway	no FORM	4.979.954	19/11/2011	SI	
Finland	no FORM	5.375.276	31/12/2010	SI	
Sweden	no FORM	9.482.855	31/12/2011	NO	
Austria	no FORM	8.401.940	31/10/2011	SI	
Slovenia	no FORM	2.050.189	01/01/2011	NO	
Netherlands	no FORM	16.655.799	01/01/2011	NO	COMBINED
Spain	FORM	46.815.910	01/11/2011	SI	
Germany	FORM	80.219.695	09/05/2011	SI	
Israel*	FORM	7.412.200	27/12/2008	SI	
United Kindom**	FORM	63.182.180	27/03/2011	SI	TRADITIONAL
Italy	FORM	59.433.744	09/10/2011	SI	

Fonte: Elaborazione degli autori nella fase di analisi della produzione censuaria di altri Paesi

* dal sito cbs.gov.il, (dati in migliaia), Paese non presente in *Census Hub*.

** Inghilterra&Galles: Totale popolazione 56.075.912

¹⁴ Nell'ambito del gruppo di lavoro inter-dipartimentale "Archivi e sistema di indagini integrate per il censimento permanente della popolazione" (ARCHETIPO) con il compito di definire il disegno strategico del censimento permanente della popolazione e delle abitazioni nella prospettiva di progressiva integrazione di registri di base e indagini statistiche, il WPC2 aveva avviato lo studio su alcuni Paesi esteri per analizzare gli strumenti adottati per il conteggio della popolazione, ovvero analizzare i metodi utilizzati per definire la popolazione totale. Per il presente lavoro sono stati presi in considerazione gli stessi Paesi, oltre a Inghilterra e Galles.

¹⁵ Per i questionari del Censimento della Popolazione e delle Abitazioni del 2011 si veda il volume Istat (2017), I contenuti informativi della rilevazione, la validazione e diffusione dei dati, ALLEGATI. (<https://www.istat.it/it/archivio/189483>)

¹⁶ Paese che non diffonde dati sul *Census Hub* e che ha effettuato il Censimento nel 2008.

Come si può notare, i Paesi con *register-based census*¹⁷ hanno una popolazione totale che varia dai 2 milioni (Slovenia) a quasi 17 milioni (Paesi Bassi). I Paesi più grandi, con una popolazione che si avvicina alla popolazione totale italiana, hanno optato per un censimento combinato o tradizionale.

I Paesi che adottano il *register-based census* hanno una popolazione totale contenuta. Probabilmente questo facilita la gestione degli archivi e, di conseguenza, l'integrazione tra le fonti amministrative. L'Italia sembra essere più correttamente confrontabile con Spagna, Germania e Inghilterra e Galles, ovvero con Paesi simili in termini di consistenza demografica, tenuta degli archivi, impostazione del censimento e diffusione censuaria.

In base a quanto emerso da una prima analisi sugli approcci censuari e sul tipo di diffusione, sono stati individuati diversi possibili scenari di diffusione futura in Italia. Nel primo scenario, "*Nessuna diffusione censuaria*", si fanno rientrare le statistiche censuarie nelle statistiche correnti tematiche, come fanno Danimarca, Svezia, Paesi Bassi, senza prevedere una diffusione *ad hoc* dei risultati censuari. Con il secondo scenario, "*Ridefinizione della diffusione in funzione di quanto c'è negli archivi*", si prevede una rottura/interruzione delle serie storiche, come fa la Slovenia. Si comunica agli utenti, nella massima trasparenza, cosa non si può più produrre e le variazioni riguardanti definizioni, universi di riferimento, classificazioni e si predispongono un piano di diffusione che massimizzi l'uso dei dati disponibili negli archivi, senza più garantire la confrontabilità con il passato. E infine il terzo scenario prevede la "*Stessa impostazione di diffusione del censimento 2011*". In questo caso si possono considerare due sotto-scenari. Nel primo, teorico ovvero poco probabile nella realtà italiana, si può garantire la stessa diffusione prodotta nel 2011 perché gli archivi esistenti, si ipotizza, contengono dati che hanno perfetta corrispondenza con quelli utilizzati ai censimenti relativamente alle definizioni, agli universi di riferimento e alle classificazioni. Nel secondo sotto-scenario, più realizzabile nel nostro paese, si predispongono un piano di diffusione in linea con i censimenti passati. I risultati censuari si desumono da archivi e indagini, sempre che la qualità sia accettabile, come accadeva in passato nei Paesi Nordici. Le altre informazioni censuarie non contenute negli archivi (o comunque di qualità non accettabile), e non rilevate da altre indagini, si raccolgono attraverso una rilevazione con questionario su un campione e/o su sottopopolazioni non coperte in altro modo, così come effettuano Spagna e Germania; in altre parole, si adotta una tecnica di *censimento combinato*, avendo definito *a priori* il piano di diffusione da produrre.

Gli scenari presentati prevedono tutti l'utilizzo di dati da fonte amministrativa e vanno nella direzione che l'Istituto intende dare alle statistiche ufficiali future: un'impostazione basata su un uso intensivo delle fonti disponibili. Cambiano i livelli di "sfruttamento" delle fonti amministrative per la diffusione in generale, e per la produzione censuaria in particolare.

È stata poi analizzata nel dettaglio la diffusione solo di quei Paesi che producono output censuario nazionale in aggiunta alla diffusione europea sul *Census Hub* (Norvegia, Finlandia, Austria, Spagna, Germania, Regno Unito, oltre a Israele¹⁸).

Numerosi e differenti sono stati i modi di approcciarsi alla diffusione censuaria da parte dei Paesi analizzati. Le differenze riguardano i contenuti trattati e diffusi in ambito censuario, le classificazioni, più o meno dettagliate in base all'argomento, oltre a nuove classificazioni ottenute combinando più informazioni (come risposte a quesiti presenti nel questionario di censimento o derivate da variabili contenute in archivi/registri) e, infine, anche le modalità con cui i risultati censuari sono stati messi a disposizione degli utenti finali. Sono state riscontrate, quindi, differenze rispetto alla diffusione italiana sia nel contenuto che nella forma (strategia e strumenti di diffusione).

Dal lato dei contenuti, escludendo informazioni che tradizionalmente non sono rilevate in Italia (religione, lingua, etnia, salute), con la ricognizione sono state individuate poche nuove tematiche/variabili¹⁹: parcheggi per edificio, piani sotto terra, piano abitazione per famiglia.

¹⁷ Si includono anche i Paesi Bassi che sono al confine tra *register-based* e *combined census*.

¹⁸ Si ricorda che l'Israele non ha pubblicato risultati sul *Census Hub* e che le pubblicazioni prese in considerazione in questo lavoro si riferiscono al censimento del 2008.

¹⁹ Alcuni Paesi diffondono informazioni sui residenti temporanei: esperimento tentato in Italia al censimento 2001 ma senza successo.

Riassumendo, sono stati individuati i seguenti elementi:

- ✓ nuove classificazioni (ottenute come combinazioni di quesiti rilevati anche in Italia) e classificazioni leggermente diverse da quelle utilizzate per la diffusione italiana;
- ✓ nuovi indicatori che possono sintetizzare alcuni fenomeni: *Deprivation dimensions*, *Approximated social grade*, *Lifestage*, *Indicators of housing standard* e *Facilities of the building* (entrambi relativi agli edifici), *Equipment standard* (relativo alle abitazioni);
- ✓ tavole che incrociano classificazioni nuove o già esistenti ma non incluse nel piano di diffusione italiano del 2011;
- ✓ focus che potrebbero essere prodotti: popolazione over 65, figli, origine-destinazione per le migrazioni, popolazione al luogo di lavoro.

In tutti i casi, si tratta di informazioni già tradizionalmente presenti nel piano di rilevazione italiano che si possono aggregare e diffondere in modo diverso. Non si è in presenza di novità di contenuto ma di nuovi incroci che potrebbero essere inseriti nel piano di diffusione dei censimenti futuri e/o di variabili e classificazioni che potrebbero essere analizzati nell'ambito della sostituibilità (da altre indagini o da fonti amministrative).

Si segnala che l'offerta informativa italiana non ha pari tra i Paesi esteri analizzati (fatta eccezione per Inghilterra e Galles) per contenuto, dettaglio classificatorio e dettaglio territoriale.

Le informazioni nuove emerse nel corso della ricognizione riguardano la parte relativa ad alloggi ed edifici, alcune delle quali probabilmente contenute nel catasto e che in futuro, quando sarà acquisito dall'Istat, potranno essere prese in considerazione per valutare il livello di qualità dei dati contenuti ai fini del loro utilizzo in ambito censuario, nella prospettiva di un ampliamento dell'offerta informativa. Sugli altri temi non si evidenzia niente di più rispetto ai temi tradizionalmente rilevati al censimento, considerati come prettamente censuari e che soddisfano una delle caratteristiche fondamentali del censimento, l'universalità.

3. Sezioni e possibili sviluppi

Le diverse fonti a disposizione, presentate nel precedente paragrafo, hanno permesso di raccogliere le informazioni necessarie per procedere alla formalizzazione delle proposte. Il secondo passo è consistito nell'analizzare le fonti per argomenti. In particolare, per il presente documento, si è proceduto facendo riferimento alle sezioni utilizzate nel questionario del Censimento 2011²⁰.

3.1 Informazioni anagrafiche e abitazioni

La valutazione delle fonti disponibili in SIM per la sezione variabili anagrafiche e abitazioni si è strutturata in due macro fasi: nella prima si è individuato un set di archivi di "interesse" per le variabili oggetto di analisi e nella seconda si è effettuata l'analisi degli archivi individuati nella prima fase. Le analisi sono state condotte utilizzando i metadati resi disponibili dal SIM per le fonti amministrative.

Nella prima fase "filtro" si sono individuati i seguenti archivi di tipo fiscale di possibile interesse per le variabili anagrafiche e di abitazioni:

- Agenzia delle Entrate – Anagrafe delle Persone Fisiche;
- Agenzia delle Entrate – Modello Unico;
- Agenzia delle Entrate – Dichiarazioni 730;
- Ministero dell'Economia e delle Finanze – Banca dati reddituale;
- Ministero dell'Economia e delle Finanze – Modello 770 statistico;
- Agenzia delle Entrate – Anagrafe delle Persone Giuridiche (possessori Partita Iva).

²⁰ Per i questionari del Censimento della Popolazione e delle Abitazioni del 2011 si veda il volume Istat (2017) "I contenuti informativi della rilevazione, la validazione e diffusione dei dati". (<https://www.istat.it/it/archivio/189483>)

E i seguenti archivi amministrativi non di tipo fiscale-economico:

- Comuni – Liste anagrafiche comunali;
- Ministero dell'Interno – Permessi di soggiorno;
- Consodata – Archivio dei contratti di telefonia fissa – Utenza privata;
- Aci – Archivio Aci;
- Agenzia delle Entrate - Dati catastali (Catasto dei terreni).

Nella seconda fase si sono esaminati nel dettaglio i tracciati record e relativi metadati/classificazioni delle fonti individuate come di possibile interesse alla ricerca di eventuali variabili per arricchire l'offerta informativa del Censimento (per alcune fonti con tracciati particolarmente complessi ci si è avvalsi anche degli stessi modelli di dichiarazione, come il Modello 730 o il modello Unico).

L'analisi dettagliata delle fonti ha determinato le considerazioni che seguono:

- Le fonti “Dichiarazioni 730”, “Unico Persone Fisiche” e “Modello 770” nel riquadro dei familiari a carico presentano l'informazione per relazione di parentela rispetto al dichiarante, in particolare se Coniuge, Primo figlio, Figlio o Altro. Tali informazioni sono di interesse per altri WP del gruppo per provare a legare l'informazione dei familiari a carico e costruire la variabile “Relazione di parentela” da rilevare nei modelli censuari.
- La possibile acquisizione del *Quadro B Redditi dei fabbricati e altri dati* sempre delle Dichiarazioni 730 potrebbe essere utile per valutare il numero di abitazioni di proprietà ed in affitto per comune per dichiarante. Andrebbe valutata la possibilità se questa informazione sia sufficiente per individuare le famiglie in abitazione in proprietà o in affitto. In questo caso non si tratterebbe di una informazione aggiuntiva a quanto già rilevato al Censimento ma potrebbe essere analizzata in termini di sostituibilità.
- L'archivio Aci oltre a poter essere utilizzato per il conteggio del numero di automobili (sostituibilità) potrebbe fornire ulteriori informazioni non ancora prodotte con la rilevazione censuaria. Dall'analisi del tracciato record sarebbero disponibili anche i dati sull'età dell'automobile e sul tipo di impiego. Alla disponibilità di auto per famiglia si potrebbe, così, produrre anche l'informazione di età e impiego dell'automobile. Inoltre, nello stesso archivio, sono inseriti anche i dati sui motoveicoli.
- Tra le fonti da poter inserire in SIM, la superficie catastale presente nei *dati catastali* permetterebbe la distinzione tra superficie totale dell'abitazione e superficie coperta per il calcolo della Tari (Tassa Rifiuti). La superficie catastale è una superficie fittizia che prende in considerazione le superfici lorde, ossia comprensive delle murature, e nel suo calcolo entrano oltre ai vani principali ed accessori diretti dell'abitazione anche, in parte, i vani accessori indiretti (cantine, soffitte, ecc.). Per questa ragione si dovrà tener presente che l'utilizzo della superficie catastale al netto delle aree scoperte come superficie dell'abitazione non coinciderà alla superficie calpestabile rilevata nei passati censimenti.
- L'archivio Consodata sui numeri telefonici fissi delle utenze private potrebbe essere utilizzato in termini di sostituibilità mentre non sono state trovate possibili variabili aggiuntive.
- Infine, la fonte Anagrafe Tributaria delle persone giuridiche, utilizzata come base principale per la costruzione del “SIM unità economiche”, potrebbe garantire il collegamento tra l'unità di base “famiglia” e il numero di Partite Iva al livello di dettaglio territoriale desiderato. In alternativa, per rendere omogenei i diversi output dell'Istituto tale collegamento potrebbe essere più opportuno se fatto con il Registro delle Imprese Attive (ASIA).

3.2 Edifici

Per quanto riguarda gli edifici e le loro caratteristiche, le fonti amministrative, sia di tipo anagrafico e sociale sia di tipo fiscale ed economico, non sembra possano contenere informazioni utili. Nonostante ciò, è stata presa in considerazione la possibilità di utilizzare i dati sulle unità immobiliari presenti nelle dichiarazioni dei redditi. Ma, dall'esame delle fonti, le problematiche relative al loro utilizzo risultano evidenti. Infatti, si tratta di dati relativi a singole unità immobiliari che non possono essere ricondotte agli edifici che le contengono, in quanto non ci sono elementi sufficienti a stabilire la loro esatta collocazione sul territorio comunale, che consentano di trovare un eventuale collegamento tramite l'indirizzo, né si hanno a disposizione riferimenti catastali. L'unico indirizzo presente sulle dichiarazioni dei redditi è il domicilio fiscale del soggetto dichiarante. Tuttavia, non si esclude l'utilizzo di queste fonti nell'eventualità di poterle integrare, tramite il codice identificativo dei soggetti proprietari, con banche dati immobiliari come ad esempio quella relativa al catasto dei fabbricati.

L'unica fonte contenente informazioni relative allo *stock* degli edifici esistenti sembra essere, infatti, la banca dati censuaria del catasto dei fabbricati detenuta dall'Agenzia delle Entrate e che potrebbe essere acquisita dall'ISTAT a livello di microdati.

Nell'ottica di poter inserire questa nuova fonte in SIM, è stato esaminato un documento contenente una descrizione schematica delle forniture dei dati catastali censuari, attualmente disponibili per i comuni attraverso un portale dedicato.

A seconda del tipo di fornitura, si distinguono il catasto dei fabbricati, registro dei fabbricati urbani, e il catasto dei terreni, registro dei terreni agricoli e fabbricati rurali. In questo lavoro si prenderà in considerazione soltanto il catasto dei fabbricati. Inoltre, la fornitura dei dati può avvenire attraverso due diverse modalità di richiesta, chiamate "attualità" e "aggiornamento". La prima riguarda l'estrazione integrale dei dati catastali a una determinata data scelta dall'utente, la seconda si riferisce soltanto a quei dati che hanno subito modifiche, all'interno di un determinato intervallo temporale, in seguito alla registrazione di nuovi atti. L'unità elementare dei microdati catastali è rappresentata dalla singola unità immobiliare alla quale, attraverso un identificativo immobile, è possibile associare le titolarità valide alla data di estrazione (in caso di fornitura di tipo "attualità") e i soggetti a cui si riferiscono le titolarità.

Data la natura molto tecnica del documento esaminato, che non sempre è ricco di metadati dettagliati utili a interpretare il significato preciso di tutte le informazioni, si sottolinea come siano necessari alcuni approfondimenti sui contenuti di interesse, al fine di valutarne l'effettiva utilizzabilità. Innanzitutto, occorrerà verificare se le singole unità immobiliari, e di conseguenza le varie caratteristiche a esse associate, siano riconducibili agli edifici che le contengono. Inoltre, nell'ottica di utilizzare una fonte amministrativa di questo tipo per il censimento permanente degli edifici bisognerà prendere in considerazione diversi aspetti, a partire da questioni di carattere definitorio a quelle relative alla qualità dei dati e all'esaustività delle informazioni.

Fermo restando lo scopo che il gruppo di lavoro si prefigge, l'utilizzo di questa come di altre fonti amministrative per l'arricchimento dell'offerta informativa, deve comunque presupporre di poter garantire le informazioni cosiddette *core*, indispensabili a garantire l'output minimo censuario. Dopodiché, occorre valutare se le informazioni aggiuntive presentano le caratteristiche delle informazioni di interesse censuario. Come, ad esempio, il dato sulle categorie catastali, che risulta essere molto dettagliato nel descrivere le varie tipologie di immobili. Così come le informazioni relative alle titolarità che caratterizzano il bene. Ed, infine, la possibilità di integrazione con altre fonti amministrative tramite i dati sugli indirizzi o tramite gli identificativi dei soggetti. Ritornando, infatti, sulle fonti relative alle dichiarazioni dei redditi, tramite l'identificativo dei soggetti, si potrebbe considerare l'informazione relativa all'utilizzo degli immobili che comprende tre macrocategorie, abitazioni principali, altri immobili a disposizione, immobili locati. Mentre, dal lato indirizzo si potrebbe provare con altre fonti, quali il registro anagrafico della popolazione o il Censimento 2011, per associare le abitazioni alle persone residenti.

3.2.1 Simulazione di analisi comparativa tra dati catastali e dati censuari

A seguito delle valutazioni fatte, quindi, si ritiene che per il censimento degli edifici il registro catastale sia, tra quelle di tipo amministrativo, la fonte da prendere principalmente in considerazione. Al fine di simulare un confronto tra le informazioni presenti nell'archivio del catasto e i dati del censimento 2011, è stato possibile utilizzare una fornitura pervenuta dall'Agenzia delle Entrate di dati catastali relativi ai fabbricati del 2012 limitatamente a quattro Comuni. In questo caso, l'unità a cui si riferiscono i microdati è rappresentata dal fabbricato e non dalle singole unità immobiliari.

Il confronto, pertanto, è stato effettuato tra le unità rilevate dal Censimento 2011 e quelle estratte dal catasto, da un lato edifici dall'altro fabbricati, nei quattro Comuni considerati. La comparazione ha riguardato sia la numerosità complessiva delle unità di tipo residenziale presenti nell'una e nell'altra fonte, sia le relative distribuzioni in base ad alcune caratteristiche, le cui variabili di riferimento sono presenti in entrambe le fonti. La scelta di analizzare soltanto le informazioni relative all'edilizia residenziale è dovuta al fatto che gli edifici non residenziali non sono stati rilevati in modo totalitario dal censimento degli edifici, in quanto hanno costituito oggetto di rilevazione solo nelle sezioni di "centro abitato" e di "nucleo abitato", mentre non rientravano nel campo di osservazione delle sezioni di "case sparse".

Il primo confronto è stato fatto tra le unità residenziali complessivamente presenti nei due archivi per ognuno dei quattro Comuni considerati. Le differenze riscontrate vanno da un minimo del 2,5% a un massimo dell'11,3% sempre a favore dei dati censuari, tranne in un caso. Tali differenze non sembrano particolarmente rilevanti, soprattutto se valutate alla luce delle seguenti considerazioni:

1. la modalità di svolgimento del censimento degli edifici (rilevazione sul campo da parte di rilevatori comunali) ha permesso di cogliere anche fabbricati non accatastati;
2. il censimento degli edifici può essere affetto da una residuale componente di sottocopertura;
3. tra le unità residenziali contenute nei due archivi (ovvero i singoli record) non esiste un perfetto rapporto 1:1. Ciò è dovuto al fatto che i fabbricati registrati al catasto vengono identificati, generalmente, con tutte le strutture che ricadono su ogni singola "area" di terreno contraddistinta da foglio e numero. Quindi, un unico fabbricato catastale potrebbe essere composto da più corpi di fabbrica, così come per uno stesso corpo di fabbrica che insiste su più "aree" vengono considerati più fabbricati. Pertanto, non sempre a un fabbricato risultante dai registri catastali corrisponde un solo edificio rilevato al censimento, così come ad un unico edificio potrebbero corrispondere più fabbricati.

All'interno delle due fonti considerate, sono state individuate le variabili (relative alle caratteristiche principali degli edifici/fabbricati) che possono essere ritenute concettualmente analoghe o simili (Tabella 2), al fine di studiarne la distribuzione.

Tabella 2 – confronto variabili Catasto-Censimento

Catasto dei fabbricati	Censimento degli edifici
Tipologia fabbricato	Tipo di edificio
Anno costruzione	Epoca di costruzione
Piani entro terra	Presenza di piani totalmente interrati
Piani fuori terra	Numero di piani fuori terra
UIU (Unità immobiliari urbane) del gruppo A (unità abitative)	Numero totale interni

Fonte: Elaborazione degli autori sulla base dei metadati disponibili nelle due fonti

Per ciascuno dei quattro comuni sono stati considerati, il numero di fabbricati residenziali con relative unità immobiliari (UIU del gruppo A) risultanti al catasto, e gli edifici residenziali con relativo numero di interni, rilevati al Censimento, in base all'epoca di costruzione. Le epoche di costruzione sono state suddivise in classi per renderle coerenti con quelle riportate sul modello di rilevazione del Censimento.

Inoltre, sono state considerate le numerosità di fabbricati e di edifici residenziali in base rispettivamente al numero di UIU (gruppo A) e di interni, poi in base alla presenza di piani interrati e infine in base al numero di piani fuori terra. Anche in questo caso, per rendere confrontabili le distribuzioni, le modalità delle tre variabili considerate sono state ricondotte a quelle utilizzate per il censimento degli edifici.

Nell'ambito delle elaborazioni qui descritte, in cui le unità di analisi sono costituite dagli edifici/fabbricati di tipo residenziale, ulteriore interesse sarebbe quello di approfondire lo studio non soltanto a un maggior livello territoriale ma anche a livello di singole unità immobiliari. Tale approfondimento consentirebbe di avere un quadro più completo al fine di effettuare le necessarie valutazioni a supporto delle decisioni sull'utilizzo delle fonti amministrative catastali per il censimento permanente. Potendo usufruire di un più ampio set di microdati di fonte catastale (UIU), si potrebbero effettuare, inoltre, valutazioni per stabilire la possibilità di sostituire le singole variabili censuarie con alcune delle informazioni presenti negli archivi catastali.

Per quanto riguarda, invece, i fabbricati non residenziali occorre evidenziare il fatto che la fonte catastale permetterebbe di avere una copertura territoriale molto più ampia del fenomeno rispetto a quella ottenuta dal censimento. Infatti, come detto in precedenza, il censimento non rileva su tutto il territorio nazionale gli edifici non residenziali, ma solo in determinate aree. Oltre a questo, un altro notevole guadagno informativo si otterrebbe grazie alla categoria catastale dei fabbricati non residenziali, in quanto presenta un elevatissimo livello di dettaglio, consentendo di rilevare molte tipologie in più rispetto a quelle rilevate dal censimento degli edifici.

3.3 Popolazione straniera

3.3.1 Criticità del censimento della popolazione straniera

Nell'ambito del sottogruppo WPD3, come già accennato, sulla base delle sezioni di cui si compone il questionario censuario del 2011 sono stati individuati diversi sottogruppi tematici, uno dei quali riguarda la popolazione straniera. E' bene ricordare che essa, definita sulla base della variabile 'cittadinanza', costituisce una componente della popolazione totale, pertanto è trasversale a tutte le aree tematiche – da quella anagrafica a quella sull'istruzione, dall'occupazione fino alla mobilità territoriale. Nel nuovo modello censuario, il ricorso alle fonti amministrative cui attingere informazioni statistiche sulla popolazione richiede un processo di trasformazione dei dati che contempla, in prima battuta, la necessità di trasporre e conciliare definizioni e classificazioni proprie di archivi costituiti per fini amministrativi con il sistema di definizioni e classificazioni di tipo statistico censuario, così come richiesto non solo dal Regolamento europeo ma anche dai fabbisogni informativi nazionali nonché dalle esigenze di confrontabilità con i passati censimenti. In altri termini, si afferma la necessità di integrare dati di varia origine con metadati specificamente censuari.

In un'ottica di sfruttamento informativo delle fonti amministrative, occorre tener conto anche che, se la "presenza" dei cittadini italiani nei diversi archivi amministrativi è facilmente riconoscibile, per gli stranieri essa non sempre è rintracciabile, soprattutto se essi vivono in condizione di irregolarità o, semplicemente, di "transitorietà" sul territorio italiano o anche di voluta "invisibilità".

La popolazione straniera rappresenta un sottogruppo la cui rilevazione, ad ogni tornata censuaria, ha destato attenzione, tanto da mettere in atto azioni mirate al contenimento degli errori censuari cui è particolarmente esposta. Per diversi ordini di motivi, essa costituisce un contingente demografico che, più di altri, è soggetto ad errori di copertura, nello specifico di sotto enumerazione. Ciò assume maggior risalto se si rapportano queste criticità censuarie al quadro normativo nell'ambito del quale sono regolati, da un lato, il censimento della popolazione, dall'altro, la presenza straniera in Italia. Ci si riferisce, in entrambi i casi, al Regolamento anagrafico²¹ il quale, nella forma dell'art. 46, ad ogni occasione censuaria stabilisce l'obbligo, per ciascun comune italiano, di procedere al

²¹ Regolamento anagrafico della popolazione residente (Legge n.223 del 30 maggio 1989) e successivo adeguamento (Legge n. 126 del 17 luglio 2015).

confronto tra risultanze censuarie e registro anagrafico della popolazione residente. Al contempo, lo stesso Regolamento richiede alla popolazione straniera non Ue²², per poter iscriversi in anagrafe, la comprovata regolarità del soggiorno, ovvero il possesso di un regolare titolo a soggiornare in Italia. Pertanto, possono essere censiti, tra i cittadini non Ue, soltanto coloro che soggiornano regolarmente sul territorio italiano, iscritti o non iscritti in anagrafe, mentre coloro che non sono in possesso di un valido permesso di soggiorno, seppure abitualmente dimoranti, sono esclusi dal campo di osservazione censuario poiché, in fase di confronto censimento-anagrafe, oltre a non poter essere iscritti in anagrafe, incorrerebbero anche nel reato di clandestinità, previsto dalla vigente normativa italiana in materia di immigrazione (Pacchetto Sicurezza)²³.

Se ai vincoli normativi si aggiunge la maggiore precarietà abitativa, lavorativa e giuridica degli stranieri rispetto alla popolazione autoctona, le criticità nelle operazioni del loro conteggio si acutizzano ulteriormente, restituendo una fotografia censuaria parziale e non del tutto rappresentativa della presenza straniera in Italia. Inoltre, a complicare la conta contribuisce il fatto che parte dei cittadini stranieri, sia comunitari che non, seppure vivano stabilmente e regolarmente sul territorio italiano, non procedono all'iscrizione in anagrafe, andando in tal modo ad innalzare il rischio di sottocopertura censuaria. Da sottolineare che anche l'anagrafe, in quanto archivio amministrativo di maggiore interesse da cui partire per la conta censuaria, risulterebbe affetta da sottocopertura.

3.3.2 Prime proposte

Alla luce delle brevi considerazioni sopra esposte circa le criticità e le caratteristiche proprie di un censimento della popolazione straniera, l'attenzione ora si focalizza sull'obiettivo specifico del presente sottogruppo, ovvero quello di analizzare il patrimonio informativo prodotto dall'Istat mediante indagini e rilevazioni ed esplorare il potenziale, talvolta ancora poco sfruttato per tale scopo, derivante da archivi e registri amministrativi disponibili ad oggi in Istituto.

L'intento è di ampliare l'offerta informativa verificando la possibilità di attivare nuovi possibili percorsi di ricerca attraverso l'uso integrato di dati provenienti da più fonti, traducibili in variabili/temi di interesse demografico, socio-economico, con specifico riferimento alla popolazione straniera.

In questa prima fase, le attività del sottogruppo "stranieri" hanno riguardato prevalentemente la ricognizione delle rilevazioni a carattere demografico condotte in Istat, sia da fonte amministrativa che da indagini campionarie, nonché l'analisi degli archivi amministrativi acquisiti sistematicamente o in corso di acquisizione da parte dell'Istituto. Da una prima rassegna del sistema informativo prodotto finora emerge, nel corso degli anni e parallelamente all'accrescersi del fenomeno migratorio, una crescente attenzione verso la popolazione straniera. Dalla fine degli anni '90 si è andata sempre più affermando l'esigenza di inquadrare la presenza straniera nella sua complessità, rilevando la molteplicità delle dimensioni che la caratterizzano, dalla struttura demografica e familiare al profilo occupazionale, dalla condizione abitativa alla salute fino alla criminalità. Sono state avviate sempre più indagini ad hoc sulla popolazione straniera e l'introduzione della variabile "cittadinanza" in tutte le rilevazioni ha consentito di raccogliere informazioni sugli stranieri anche nell'ambito di indagini non specificamente dedicate.

Nel corso dei decenni, anche il censimento della popolazione ha compiuto lo sforzo di adeguamento alla realtà in trasformazione, progettando e mettendo in atto interventi mirati per meglio cogliere la componente straniera, sia in un'ottica di innalzamento della sua copertura censuaria che di approfondimento conoscitivo del fenomeno. Sono stati introdotti nuovi quesiti, utili alla individuazione di specifiche sottopopolazioni di stranieri, per consentire di inquadrare il fenomeno migratorio nella sua complessità e di individuare anche il background migratorio. Inoltre, l'eshaustività, la ricchezza tematica e il fine dettaglio territoriale al quale sono riferiti i dati censuari fanno del censimento una preziosa miniera di informazioni che tuttavia, essendo di *stock*, invecchiano velocemente e non

²² Gli stranieri provenienti dagli Stati membri Ue non necessitano di permesso/carta per soggiornare in Italia.

²³ Legge n. 94 del 15 luglio 2009.

riescono a cogliere correntemente la dinamica dei fenomeni, il loro cambiamento nel breve periodo. Ma il censimento permanente dovrebbe ovviare a questo limite censuario attraverso gli aggiornamenti dei dati di *stock* con i dati di flusso o mediante l'impiego di dati di *stock* a cadenza più ravvicinata, permettendo di passare da una cadenza decennale ad una annuale. Inoltre, lo sfruttamento di dati longitudinali potrà conferire alla rilevazione censuaria un valore aggiunto, funzionale allo studio dei fenomeni sociali e demografici in continuo divenire. Ciò è tanto più vero per una popolazione di recente insediamento sul territorio italiano, quindi ancora in fase di assestamento.

È innegabile che l'ampiezza del ventaglio informativo è in genere inversamente proporzionale alla finezza del dettaglio territoriale, pertanto l'impiego massiccio di dati amministrativi, integrati con dati campionari, richiede un'attenta riflessione sul rischio di compromettere sia il dettaglio territoriale che quello informativo. A titolo esemplificativo, le rilevazioni demografiche, di fonte amministrativa, specifiche sulla popolazione straniera, come Strasa²⁴ o P3²⁵ rilevano e diffondono soltanto dati sulla struttura demografica fino a livello comunale, mentre gran parte delle indagini campionarie rilasciano i dati demografici, socio-economici, sanitari, giudiziari ecc. ad un dettaglio territoriale meno fine, quello regionale. Con particolare riferimento al Portale Istat sugli stranieri²⁶ e al data *warehouse* Immigrati.Stat²⁷, sono state analizzate le indagini/rilevazioni effettuate in Istituto specificamente sulla popolazione straniera o, genericamente, sulla popolazione totale ma contenenti l'informazione sulla cittadinanza degli individui. La ricognizione è stata realizzata esaminando le indagini in un'ottica di una loro utilizzabilità rispetto agli obiettivi prefissi, ovvero rispetto alla possibilità di integrare informazioni provenienti da diverse fonti e di individuarne di nuove, in una prospettiva di ampliamento dell'offerta informativa, tenendo conto tuttavia dell'esigenza di scremare la sovrapproduzione di dati. L'analisi -che ha prodotto un quadro sinottico dettagliato- è stata condotta sulle seguenti caratteristiche: la tipologia, ovvero se i dati derivano da un'indagine campionaria o da fonte amministrativa o da censimento; la specificità o meno, ovvero se si tratta di indagine ad hoc sugli stranieri oppure sull'intera popolazione; la denominazione dell'ente fornitore di dati e dell'ente per il trattamento e la diffusione dei dati; il dettaglio territoriale di diffusione dei dati; le variabili disponibili nel modello di raccolta dei dati e le variabili rilevate e diffuse; la periodicità dell'indagine/rilevazione; l'universo di riferimento, in particolare quali stranieri vengono conteggiati e rilevati e quali sfuggono alla rilevazione; la rilevanza ai fini del solo conteggio di popolazione o anche dei contenuti informativi; le principali analisi consentite; l'appartenenza o meno dell'indagine nel Programma Statistico Nazionale.

Dopo aver esaminato le indagini statistiche, anche quelle che già sfruttano alcuni archivi amministrativi, si è proceduto ad una ricognizione delle fonti amministrative disponibili in Istat (per lo più presenti nel sistema SIM) per lo studio della popolazione straniera, sia in termini di conteggio sia di caratteristiche, soprattutto in un'ottica di potenzialità delle fonti amministrative a tale scopo²⁸. Oltre ad aver individuato gli "archivi benchmark" (ossia gli archivi relativi all'anagrafe della popolazione di fonte fiscale e comunale) utili a identificare la popolazione complessiva presente sul territorio nazionale, sono stati individuati gli "archivi contenenti la variabile cittadinanza", fondamentali per la definizione della popolazione straniera oltre che per l'arricchimento informativo di tale sottoinsieme della popolazione. Successivamente sono stati esaminati quegli "archivi contenenti variabili appartenenti ad aree tematiche rilevanti (Lavoro, Istruzione, Redditi)" che, seppur non inclusivi della variabile cittadinanza ma comunque contenenti dati individuali, rappresentano degli archivi fondamentali per caratterizzare la popolazione straniera secondo informazioni di rilievo.

La disponibilità di un codice identificativo univoco, anonimo e stabile nel tempo ("codice individuo"), attribuito dal sistema SIM, permette di riferire alla popolazione straniera, definita a partire dagli "archivi contenenti la variabile cittadinanza", le informazioni desumibili dagli "archivi conte-

²⁴ Popolazione residente comunale straniera per sesso ed anno di nascita, <http://demo.istat.it/strasa2016/index.html>

²⁵ Movimento e calcolo della popolazione straniera residente, <http://demo.istat.it/str2015/index.html>

²⁶ <http://www.istat.it/it/immigrati>

²⁷ <http://stra-dati.istat.it/>.

²⁸ Si intende precisare che anche per tale ricognizione è stato prodotto un quadro sinottico.

nenti variabili appartenenti ad aree tematiche rilevanti”. Ciò consente di arricchire il contenuto informativo del Censimento anche rispetto ad aree tematiche mai esaminate sinora, quali ad esempio l’area dei dati fiscali.

Nell’ambito della ricognizione delle fonti amministrative, sono state definite alcune meta-informazioni utili a valutare l’impiego delle fonti stesse: innanzitutto l’area tematica principale di appartenenza (“Stranieri” e “Non Stranieri”, per intendere rispettivamente le fonti specifiche per la sottopopolazione in esame e le fonti relative alla popolazione complessiva ma inerenti tematiche rilevanti quali: Lavoro, Istruzione, Redditi); la presenza della variabile ‘cittadinanza’, utile per la corretta individuazione della popolazione straniera; la presenza delle variabili connesse al ‘paese di nascita’ per una lettura congiunta delle informazioni che tenga conto dei due criteri (cittadinanza e paese di nascita); la presenza di variabili di rilievo distinte tra “*core*” e “non *core*” rispetto al Censimento; una variabile ‘utilizzo’ per dar conto dell’utilità dell’archivio ai fini del conteggio e/o del contenuto; il ‘dettaglio territoriale’; infine la presenza negli attuali sistemi SIM e l’anno di riferimento ultimo disponibile dei dati.

La presente ricognizione ha permesso di individuare un elevato numero di “archivi contenenti la variabile cittadinanza”, tra questi: l’archivio di stato civile comunale che coglie le nascite, le morti e i matrimoni, gli archivi del Ministero degli interni relativi ai permessi di soggiorno e alle acquisizioni di cittadinanza italiana; gli archivi Inps riferiti ai lavoratori extracomunitari e ai lavoratori domestici; quelli dell’Inail inerenti i lavoratori parasubordinati e i lavoratori interinali; quelli del Ministero dell’istruzione relativi all’anagrafe degli studenti; e infine quelli delle CCIAA riferiti ai soci delle società di persone e società di capitale. Sebbene non sia ancora disponibile in Istat e nel SIM, vale la pena ricordare anche la fonte delle Comunicazioni Obbligatorie del Ministero del Lavoro. Alle fonti più specifiche della sola area Stranieri si aggiungono dunque quelle che tengono conto congiuntamente delle aree tematiche Stranieri/Lavoro, Stranieri/Istruzione; Stranieri/Imprese.

Come precedentemente detto, tra gli archivi utili allo studio della popolazione straniera in un’ottica di estensione del contenuto informativo del Censimento, sono stati anche proposti quelli “contenenti variabili appartenenti ad aree tematiche rilevanti” e in particolare nell’area “Lavoro” sono state individuate le seguenti fonti: gli archivi dell’Inps sulle denunce contributive dei lavoratori dipendenti (EMens), sui beneficiari della Cig a pagamento diretto, sulla dichiarazione di manodopera agricola, sui lavoratori autonomi in agricoltura, sui lavoratori parasubordinati, sui lavoratori artigiani e commercianti. Nell’area tematica “Istruzione”, gli archivi del Ministero dell’istruzione relativi agli iscritti e iscrizioni universitarie, al personale universitario, ai laureati, al personale delle scuole statali, e infine agli esiti scolastici. E infine nell’area tematica “Redditi” sono state focalizzate le seguenti fonti: gli archivi dell’Ag. Entrate sulle dichiarazioni 730, 770, Unico Persone Fisiche.

L’elevato numero di fonti e di variabili che ciascuna fonte racchiude fornisce una prima indicazione di massima dell’enorme potenziale informativo che le fonti amministrative possono offrire a integrazione e sostituzione di alcuni dei quesiti del Censimento per la popolazione straniera, soprattutto quando le variabili vengono esaminate secondo i criteri proposti (core/no core, conteggio/contenuto, riferimento temporale, dettaglio territoriale).

4. Ulteriori proposte di approfondimento

Nel precedente paragrafo 3 sono stati evidenziati i possibili sviluppi di specifiche sezioni informative del censimento e, in maniera trasversale, i risultati delle analisi svolte sulla popolazione straniera. Oltre l’individuazione di singole variabili aggiuntive rispetto al tradizionale output censuario, è possibile formalizzare delle proposte valutando anche la possibilità di sfruttare le informazioni dei sistemi relazionali sviluppati nel SIM, l’esperienza che l’Istituto ha acquisito con il progetto Archimede e la diffusione di nuovi incroci di variabili in grado di offrire ulteriori chiavi di lettura dei fenomeni indagati.

La novità sostanziale riguardo l’ampliamento informativo e lo sviluppo di nuove aree di ricerca per il censimento è rappresentata dalla possibilità di disporre di microdati utili per analisi di tipo longitudinale con un elevatissimo grado di dettaglio.

La disponibilità di fonti amministrative, Registri statistici integrati, output annuali e microdati consente, in linea teorica, differenti modalità di diffusione delle informazioni. Sulla base di questa considerazione, la possibilità di formalizzare delle proposte di nuove informazioni da diffondere con il censimento permanente deve tener conto di alcune scelte a monte riguardo: le fonti da utilizzare, le popolazioni da indagare e le modalità con cui diffondere le variabili.

Qui di seguito sono illustrati due esempi che mostrano come la possibilità di formalizzare le proposte sia flessibile e adattabile ai singoli fenomeni oggetto di analisi. Inoltre, la modalità proposta, permette di tener conto delle scelte strategiche fatte a monte circa l'insieme di caratteristiche della popolazione da diffondere con il Censimento.

Come indicato nel paragrafo 3.1, l'insieme delle fonti e dei Registri statistici sulle unità economiche potrebbero essere utilizzati per collegare gli individui, o le famiglie, a dati di tipo economico. Oltre l'indicazione sulla presenza di individui con Partita Iva all'interno delle famiglie, si potrebbe cogliere il ruolo societario svolto dai singoli componenti delle stesse. In figura 4, sono evidenziate, nel lato sinistro, le diverse combinazioni informative ottenibili sulla base di fonti, popolazioni interessate e variabili e, nel lato destro, le soluzioni che, ad oggi, appaiono preferibili (in grigio).

Figura 4 – Proposta: ruolo societario

Argomento	Fonte	Popolazione	Variabili	Argomento	Fonte	Popolazione	Variabili
RUOLO SOCIETARIO	hp I: solo DA	hp A: tutti i potenziali soggetti	hp 1: flag si/no hp 2: tipo ruolo	RUOLO SOCIETARIO	hp I: solo DA	hp A: tutti i potenziali soggetti	hp 1: flag si/no hp 2: tipo ruolo
	hp II: solo RS	hp B: solo donne			hp II: solo RS	hp B: solo donne	
	hp III: DA + RS	hp C: solo stranieri			hp III: DA + RS	hp C: solo stranieri	
		hp D: ...				hp D: ...	

Fonte: Elaborazione degli autori – formalizzazione delle proposte sulla base dei microdati potenzialmente disponibili

In primo luogo, l'informazione sul ruolo societario, all'interno dell'Istituto, è presente come dato d'origine nei file di provenienza CCIAA (DA) e come output statistico (RS). Oltre l'intero universo dei soggetti coinvolti, potrebbero essere realizzati dei focus specifici per sottopopolazioni, ad esempio donne e stranieri. Infine, l'informazione aggiunta al Censimento potrebbe essere un semplice *flag* di presenza/assenza del fenomeno oppure potrebbero essere definiti i singoli ruoli. In termini di fattibilità, ad oggi, sembra preferibile partire da output statistici dell'Istituto e considerare delle sottopopolazioni specifiche in modo da caratterizzare maggiormente il dato censuario rispetto alle fonti di origine. In questo caso si potrebbe valutare l'ipotesi di considerare la famiglia, e non il singolo individuo, come unità di analisi. Si rileva, inoltre, che il legame con il Registro Asia potrebbe anche permettere un collegamento dell'informazione alle sole imprese attive nell'anno. Infine, per il Censimento, un semplice segnale di presenza potrebbe essere sufficiente per una prima analisi del fenomeno rimandando alle rilevazioni tematiche il compito di integrare il set di informazioni disponibili.

Considerare delle sottopopolazioni potrebbe essere una valida soluzione per risolvere situazioni in cui il patrimonio informativo delle fonti amministrative non risulti essere completo. Riguardo i titoli di studio, ad esempio, lo sviluppo del Registro Titoli di studio all'interno dell'Istituto, il progetto Archimede e le analisi specifiche svolte per il Censimento hanno evidenziato come ci siano delle fonti amministrative attualmente non in possesso dell'Istituto. In figura 5 sono evidenziate alcune possibilità di analisi.

Figura 5 – Proposta: minori dopo la scuola secondaria di primo grado

Argomento	Fonte	Popolazione	Variabili	Argomento	Fonte	Popolazione	Variabili
MINORI: cosa fanno dopo la scuola secondaria di primo grado	hp I: solo DA	hp A: tutti i potenziali soggetti	hp 1: flag sì/no hp 2: aspetto longitudinale	MINORI	hp I: solo DA	hp A: tutti i potenziali soggetti	hp 1: flag sì/no hp 2: aspetto longitudinale
	hp II: solo RS	hp B: solo dopo scuola primaria			hp II: solo RS	hp B: solo dopo scuola primaria	
	hp III: DA + RS	hp C: solo dopo scuole "medie"			hp III: DA + RS	hp C: solo dopo scuole "medie"	
	hp D: ...	hp D: ...			hp D: ...	hp D: ...	

Fonte: Elaborazione degli autori - formalizzazione delle proposte sulla base dei microdati potenzialmente disponibili

Un focus specifico per seguire nel tempo i minori dopo la scuola secondaria di primo grado potrebbe generare delle informazioni utili per vari aspetti. Il censimento permanente nel suo aspetto longitudinale potrebbe contenere le prime indicazioni in termini di segnali e sottopopolazioni da cui partire per seguire fenomeni come l’abbandono scolastico e l’inserimento lavorativo dei minori. In combinazione con le indagini (di copertura del censimento e sociali) potrebbe garantire anche la stima di ambiti non colti attualmente dalle fonti amministrative. In termini predittivi potrebbe essere utile per i comuni per programmare le politiche sui trasporti locali. Chiaramente, lo sviluppo di output statistici specifici, ad esempio il progetto Archimede “Percorsi di istruzione, formazione e inserimento lavorativo” (vedi paragrafo 2.2), consentirebbe di seguire nel tempo popolazioni più ampie e il focus sui minori sarebbe un *di cui* dell’output completo.

Dall’esperienza tratta nell’ambito del gruppo di lavoro, la modalità di formalizzazione delle proposte di contenuti informativi nuovi per il censimento permanente ritenuta maggiormente opportuna è quella che prevede l’uso dei Registri statistici integrati con il necessario supporto di rilevazioni specifiche. Considerando l’integrazione con i settori tematici, il Censimento, in alcuni casi, potrebbe contenere dei semplici segnali di presenza/assenza di particolari fenomeni analizzati in altri ambiti ottenendo coerenza tra gli output e possibilità di ricerca a diversi livelli di dettaglio.

5. Criticità e prospettive

Nel presente report sono state illustrate le decisioni operative e metodologiche adottate nell’ambito del WP D3 per identificare potenziali contenuti informativi nuovi da diffondere con il censimento permanente.

La prima criticità riscontrata è stata quella di capire cosa si potesse intendere per nuovo contenuto informativo del censimento. Gli ambiti socio-economici da poter cogliere con l’uso combinato di fonti amministrative e indagini mirate sul campo sono, infatti, molteplici. Per questo motivo si è scelto di valutare, prima di tutto, come ampliare le informazioni sulle sezioni esistenti del Censimento 2011.

La valutazione svolta per la sezione “variabili individuali” ha evidenziato come molte delle fonti disponibili nel SIM siano utili soprattutto in termini di sostituibilità. Alcuni spunti interessanti per ampliare l’offerta informativa potrebbero riguardare: la disponibilità di mezzi di trasporto propri; il collegamento tra individui, famiglie e alcune informazioni disponibili dal lato delle tradizionali statistiche economiche.

Per la valutazione della sezione “abitazioni” sono state considerate anche fonti acquisite, ad oggi, parzialmente dall’Istituto o ancora da acquisire. Potenzialmente, tali dati potrebbero essere utili per valutare il numero di abitazioni di proprietà ed in affitto per comune e per dichiarante e a sostituire l’informazione della superficie calpestabile con la superficie catastale che permette di individuare anche la quota di superficie delle aree scoperte dell’abitazione come balconi o terrazzi.

Un'altra indicazione che potrebbe arricchire il patrimonio informativo censuario, desumibile dai fogli catastali, è il "piano dell'abitazione".

Per gli "edifici" l'unica fonte contenente informazioni relative allo *stock* degli edifici esistenti sembra essere la banca dati censuaria del catasto dei fabbricati (Agenzia delle Entrate) per la quale bisognerà considerare la compatibilità tra le informazioni *core* del censimento e i dati amministrativi in termini di definizioni, qualità dei dati e esaustività delle informazioni. Tra le informazioni aggiuntive potrebbero essere valutate, ad esempio, il dato sulle categorie catastali, che risulta essere molto dettagliato nel descrivere le varie tipologie di immobili e le informazioni relative alle titolarità che caratterizzano il bene.

Relativamente al tema 'Stranieri' il valore aggiunto derivante dalla longitudinalità del censimento permanente e il fine dettaglio territoriale, potrebbero consentire, in prospettiva, la costituzione di osservatori volti a seguire, nel tempo, sottogruppi di interesse o fenomeni sociali emergenti, sfruttando l'approccio integrato delle diverse fonti di dati disponibili e consolidando indagini già esistenti o incorporandone altre, con l'obiettivo di prevenire o, quantomeno, di ridurre al minimo la sovrapproduzione statistica. Si pensi alla possibilità di assumere come oggetto di indagine un sottogruppo di stranieri, quali ad esempio gli stranieri nati in Italia, e seguirlo a cadenza annuale attraverso l'integrazione di dati di flusso amministrativi, provenienti da più fonti e i dati censuari raccolti annualmente. Oppure, ribaltando il punto di osservazione, assumere il fenomeno di interesse come oggetto di indagine, quale ad esempio la segregazione spaziale delle diverse collettività straniere e studiarne i modelli insediativi e i cambiamenti nel tempo e nello spazio. D'altro canto, anche l'ottica di genere rappresenta un'angolatura da cui osservare il fenomeno migratorio nelle sue forme di integrazione sociale, lavorativa, culturale, economica.

Infine, un ulteriore elemento innovativo rispetto al passato, che attiene più alla "conta" che al "contenuto", potrebbe risiedere nella possibilità di stimare/conteggiare gli stranieri irregolari. Sebbene risulti impresa ardua e suscettibile di critiche, rappresenterebbe però un passo evolutivo nella statistica ufficiale, che va nella direzione di un'armonizzazione con altri paesi europei e, soprattutto, di un riconoscimento oggettivo della presenza di individui che vivono sul territorio italiano, senza tema di errori o giudizi.

Sulla base delle analisi svolte in termini di metadati, possono essere tratte alcune indicazioni. In linea generale, potrebbe essere preferibile inserire, ove possibile, nuove variabili in collegamento con Registri Statistici dell'Istituto o, comunque, basi informative comuni per diversi processi statistici. In secondo luogo, il censimento permanente potrebbe contenere dei segnali di presenza di particolari fenomeni colti, con maggiore dettaglio, nei rispettivi ambiti tematici. Infine, sembra evidente come la principale novità del censimento permanente sia rappresentata dall'output di tipo annuale che permette di svolgere analisi di tipo longitudinale. Da considerare, però, che tale opportunità dovrà essere accompagnata da un adeguato sviluppo della strategia di diffusione prevedendo anche la realizzazione di output statistici di tipo diverso rispetto a quelli tradizionali.

La modalità proposta è estendibile anche ad altre linee di ricerca che potrebbero emergere nel proseguimento delle attività relative alla realizzazione del censimento permanente della popolazione e delle abitazioni. Da sottolineare nuovamente come le indicazioni presenti in questo documento derivino dalla sola analisi dei metadati disponibili e che la valutazione definitiva sulla possibilità di aggiungere contenuti informativi nuovi di carattere censuario richiederà ulteriori analisi da svolgere sui microdati e lo sviluppo di specifiche e adeguate metodologie di stima.

Riferimenti bibliografici

- Ambroselli, S., Di Bella G., (2014). *Towards a more efficient system of administrative data management and quality evaluation to support statistics production in Istat*. Paper presented to the European Conference on Quality in Official Statistics Q2014. Vienna: 2-5 June.
- Ambroselli, S. 2015. *I codici identificativi univoci all'interno del SIM (Sistema Integrato di Microdati)*. Roma: Istat Working Papers 5 2015.
- Ambroselli, S, Garofalo, G. (2015). *Reversing the flow: from an integrated system of administrative microdata to an infrastructure for the users*. Session 6C -Enterprise architecture and integration. NTTS 2015. Bruxelles: Marzo 2015.
- Bakker, B.F.M. 2011. Micro-integration: State of the Art. ESSnet on Data Integration WP1 State of the Art on Statistical Methodologies for Data Integration.
- Calzaroni M., Congia M.C., Garofalo G., Righi A., Rizzi R., Tronti L. “Using administrative sources for studying rare phenomena: foreign female employment in Italy”. In International Working Party on Labour Market Segmentation (IWPLMS)- 26th Annual Conference, Berlino, 8-11 Settembre 2005 (pubblicato sul sito <http://www.dynamoproject.eu/conference.shtml>).
- Cassani B., Giudici C., Rizzi R. “Immigrant entrepreneurship through the economic crisis in Italy”- presentato alla SIS XLVI Riunione scientifica, Roma, 20-22 giugno 2012. Pubblicato in Società Italiana di Statistica. Atti della XLVI Riunione scientifica. P. 78.
- Cortese A., Gallo G., Paluzzi E.. “Il censimento della popolazione straniera: opinioni a confronto sul principale aspetto definitorio”. In “Contributi Istat, n.1/2010, Roma, gennaio 2010
- Eurostat, CENSUS HUB. <https://ec.europa.eu/CensusHub2>
- Garofalo, G., A cura di, 2014. *Il Progetto ARCHIMEDE Obiettivi e risultati sperimentali*. Roma: Istat Working Papers N. 9/2014.
- IDOS Centro Studi e Ricerche (a cura di). “Dossier Statistico Immigrazione 2015”, Roma, ottobre 2015.
- IDOS Centro studi e Ricerche (a cura di). “Rapporto Immigrazione e Imprenditoria 2015”, Roma, novembre 2015.
- Istat (2011). 15° Censimento della popolazione: i risultati. <http://dati-censimentopopolazione.istat.it/Index.aspx>
- Istat (2017). *I contenuti informativi della rilevazione, la validazione e diffusione dei dati*. Atti del 15° Censimento generale della popolazione e delle abitazioni. Roma: Istat. (<https://www.istat.it/it/archivio/189483>)
- Regulation (EC) No 763/2008 of the European Parliament and of the Council of 9 July 2008 on population and housing censuses.*
- Regulation (EC) No 1201/2009 of the European Parliament and of the Council of 30 November 2009 implementing Regulation (EC) No 763/2008 on population and housing censuses as regards the technical specifications of the topics and of their breakdowns.*
- Regulation (EU) No 519/2010 of the European Parliament and of the Council of 16 June 2010 adopting the programme of the statistical data and of the metadata for population and housing censuses provided for by Regulation (EC) No 763/2008 of the European Parliament and of the Council.*
- Runci M., C., Di Bella G., Galiè L. (2016). Il sistema di integrazione dei dati amministrativi in Istat. Istat Working Papers 18 2016.
- UNECE (2007). *Register-based statistics in the Nordic countries: review of best practices with focus on population and social statistics*. United Nations Publication, ISBN 978-92-1-116963-8.

- UNECE (2011). *Using Administrative and Secondary Sources for Official Statistics - A Handbook of Principles and Practices*. United Nations Publication.
- Wallgren A., Wallgren B. (2011). *To understand the Possibilities of Administrative Data you must change your Statistical Paradigm! Proceedings of the Survey Research Methods Section*. American Statistical Association, Invited Papers.
- Zhang, Li-Chun (2012). *Topics of statistical theory for register-based statistics and data Integration*. *Statistica Neerlandica*66, (1), 41-63.

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, ecc.).

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.