

# istat working papers



N. 9  
2014

## **Il Progetto ARCHIMEDE obiettivi e risultati sperimentali**

*Giuseppe Garofalo*



# istat working papers

N. 9  
2014

## **Il Progetto ARCHIMEDE obiettivi e risultati sperimentali**

*Giuseppe Garofalo*

### **Comitato scientifico**

Giorgio Alleva  
Tommaso Di Fonzo  
Fabrizio Onida

Emanuele Baldacci  
Andrea Mancini  
Linda Laura Sabbadini

Francesco Billari  
Roberto Monducci  
Antonio Schizzerotto

### **Comitato di redazione**

Alessandro Brunetti  
Romina Fraboni  
Maria Pia Sorvillo

Patrizia Cacioli  
Stefania Rossetti

Marco Fortini  
Daniela Rossi

### **Segreteria tecnica**

Daniela De Luca   Laura Peci   Marinella Pepe   Gilda Sonetti

## **Istat Working Papers**

Il Progetto ARCHIMEDE  
obiettivi e risultati sperimentali

N. 9/2014

ISBN 978-88-458-1819-6

© 2014

Istituto nazionale di statistica  
Via Cesare Balbo, 16 – Roma

Salvo diversa indicazione la riproduzione è libera,  
a condizione che venga citata la fonte.

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

## Il Progetto ARCHIMEDE obiettivi e risultati sperimentali

Giuseppe Garofalo

### Sommario

*Il Progetto ARCHivio Integrato di Microdati Economici e DEMografici (ARCHIMEDE) ha 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) 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 deve essere raggiunto attraverso lo sfruttamento dei contenuti informativi delle basi di dati amministrative integrate nel Sistema integrato dei microdati (Sim). Nel corso del 2013 anno sono state progettate e condotte tre sperimentazioni relativamente ai temi "Popolazione insistente" (identificazione, classificazione e quantificazione della popolazione che usa il territorio), "Precarietà lavorativa" (identificazione, classificazione e qualificazione dei lavoratori con contratti di lavoro definiti precari) e "Condizioni socio-economica delle famiglie" (costruzione di una struttura informativa sulle famiglie che consenta di analizzare vari aspetti delle loro condizione socio-economica). Il fine delle sperimentazioni era 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. Il documento presenta i risultati della sperimentazione effettuata nel corso del 2013.*

**Parole chiave:** Fonti amministrative, basi di dati integrati, city users, condizioni delle famiglie, precarietà lavorativa.

### Abstract

*Istat Project ARCHIMEDE (Integrated archive of economic and demographic micro data) aims at expanding Istat information offer by producing longitudinal paths (for example, social and economic) and cross-sectional collections of micro data to be made available to users and useful to social and economic research, to sectorial and territorial planning, and to public policy evaluation at national, regional and local levels. This objective has to be achieved through the exploitation of administrative database information contents integrated into Istat platform SIM (Integrated Micro data System). During the year 2013 three experiments were designed and conducted in relation to the themes "Resident population" (identification, classification and quantification of the population using the territory), "Precarious employment" (identification, classification and qualification of workers with precarious employment contracts) and "Household socio-economic conditions" (construction of an information structure on households to analyse various aspects of their socio-economic status). The purpose of the experimentation was to assess the real project potential on the one hand, and to propose and assess the feasibility of specific statistical products and systems for the dissemination of information outputs, on the other. The paper presents the results of the experiments carried out.*

**Keywords:** Administrative sources, integrated databases, city users, household conditions, precarious employment.



## Indice

	Pag.
<b>1. Le caratteristiche generali del Progetto Archimede</b> .....	9
1.1. Il Sistema Integrato dei Microdati .....	9
1.2. Dal SIM al Progetto Archimede .....	9
1.3. Una modifica nella visione della produzione di informazioni statistiche .....	10
1.4. Le macrofunzioni del progetto .....	12
1.5. Gli Output Informativi .....	14
1.6. La tutela della riservatezza nel trattamento e diffusione dei dati .....	15
1.7. La progettazione dei sistemi di diffusione dei dati .....	15
1.8. I sotto-progetti sperimentali realizzati .....	17
<b>2. Il sotto-progetto sperimentale: popolazione insistente</b> .....	19
2.1 Le caratteristiche della sperimentazione .....	19
2.2 Il Sistema P&P .....	22
2.3 I risultati .....	28
2.4 Prime ipotesi di prodotti .....	28
2.5 Sviluppi futuri .....	34
<b>3. Il sotto-progetto sperimentale: precarietà lavorativa</b> .....	40
3.1 Le caratteristiche della sperimentazione .....	40
3.2 I risultati .....	44
3.3 Le potenzialità informative della base dei dati .....	48
3.4 Sviluppi futuri .....	56
<b>4. Il sotto-progetto sperimentale: La precarietà socio economica delle famiglie</b> .....	57
4.1 Le caratteristiche della sperimentazione .....	57
4.2 I risultati .....	61
4.3 Le potenzialità informative della base dei dati .....	69
4.4 Sviluppi futuri .....	75
<b>5. Conclusioni</b> .....	76
<b>Riferimenti bibliografici</b> .....	78





## 1. Le caratteristiche generali del Progetto Archimede

### 1.1 Il Sistema Integrato dei Microdati

Il Progetto ARCHIMEDE utilizza come base informativa di riferimento il Sistema Integrato dei Microdati (SIM). Questa è una infrastruttura statistica di base realizzata dall'Istat presso il Dipartimento dei censimenti e degli archivi amministrativi e statistici (DICA) nella quale vengono integrati dati elementari riferiti a individui, famiglie e unità economiche provenienti da una pluralità di basi dati di fonte amministrativa. Il sistema include informazioni concernenti le:

- Caratteristiche di individui e famiglie (es. occupazione, titolo di studio);
- Caratteristiche delle **unità** economiche in cui gli individui realizzano le loro attività (o la loro vita);
- Caratteristiche dei **luoghi** di residenza, lavoro, studio, ecc.;
- Tipologie di **relazioni** tra individui, unità economiche e luoghi.

Obiettivo principale del SIM è l'individuazione della presenza di una stessa unità elementare (persona fisica, unità, luogo, relazione) nelle diverse fonti e l'attribuzione ad essa di una chiave di identificazione univoca e stabile nel tempo. A partire dai singoli archivi in cui ciascuna entità rappresentata è individuata, è possibile definire la combinazione delle fonti in cui una determinata unità è presente in un intervallo di tempo stabilito attraverso procedimenti di riconoscimento (deterministici e/o probabilistici) tra entità omogenee. Si crea in tal modo un collegamento logico/fisico tra la totalità delle informazioni relative ad una specifica entità nelle diverse fonti di input.

Il Sim è quindi un sistema integrato, poiché realizza l'identificazione di ciascuna **entità** (*famiglia, individuo, unità e loro relazioni*) **del sistema** nelle diverse fonti (mediante un **ID number** unico e stabile nel tempo) definendo, per ciascuna **entità**, le relazioni logiche e fisiche (*nel tempo e nello spazio*) che intercorrono tra le informazioni ad essa riferite nelle varie fonti.

Attualmente il SIM integra circa 60 archivi gestiti da varie amministrazioni pubbliche. In maniera del tutto indicativa, di seguito, si citano i principali:

- *Archivi fiscali*: Anagrafi tributarie, Data base reddituale (integrazione di 730/Unico/770), studi di settore, dichiarazioni Iva, 770, ecc.;
- *Archivi camerali*: Registro imprese, Registro delle persone di imprese, Registro dei soci, Bilanci civilistici, ecc.;
- *Anagrafi comunali*;
- *Archivi previdenziali*: e-mens, archivi dei lavoratori autonomi (artigiani, commercianti agricoli), manodopera agricola, lavoratori domestici, archivio ex Enpals, archivio ex-Inpdap, casellario dei pensionati, archivio lavoratori parasubordinati;
- *Archivi delle prestazioni non pensionistiche* (in fase di acquisizione): assegni familiari, disoccupati, dipendenti in mobilità, Lsu;
- *Archivi assicurativi*: parasubordinati e interinali;
- *Archivi Miur*: anagrafe degli studenti, anagrafe delle scuole, anagrafe del personale docente e non delle scuole, anagrafe degli universitari, anagrafe delle università, anagrafe del personale docente e non delle università.

Il SIM si compone di sette sottosistemi, due dei quali sono riferiti alle entità (sottosistemi individui e unità economiche), due alla presenza sul territorio delle entità (sottosistemi luoghi individui e luoghi unità). Tre infine rappresentano le relazioni che intercorrono fra le entità (relazioni fra individui/individui, relazioni fra unità economiche e relazioni fra individui e unità).

### 1.2 Dal SIM al Progetto Archimede

Il SIM individua strutture di dati, riferiti a unità elementari appartenenti a specifiche popolazioni di interesse statistico, che garantiscono meglio e in maniera standardizzata (grazie ai processi di integrazione, identificazione e omogeneizzazione sviluppati) il tradizionale utilizzo di fonti amministrative per finalità statistiche, quali:

- la realizzazione e l'aggiornamento di registri statistici (un esempio è il registro ASIA);

- la realizzazione diretta di stime di variabili per domini di interesse;
- il supporto alle rilevazioni statistiche in relazione sia al miglioramento di processi di produzione (disegno campionario, stime per piccole aree), sia all'identificazione di informazione ausiliaria nella fase di controllo dei dati.

Poiché il SIM contiene un patrimonio informativo riferito non solo agli universi delle entità (*amministrative*) individuate, ma anche ad un numero di variabili (*amministrative*) incredibilmente ampio, può essere individuato come una infrastruttura da cui derivare “direttamente” informazione statistica completamente nuova anche al fine di sviluppare analisi statistiche nel tempo e per ambiti territoriali ristretti.

All'interno di quest'ultimo aspetto si colloca il progetto Archimede.<sup>1</sup>

Istituito nel marzo 2013,<sup>2</sup> il progetto ha l'obiettivo di realizzare: *l'ampliamento dell'offerta informativa dell'Istat mediante la produzione di collezioni di dati elementari di tipo longitudinale (ad es. percorsi sociali ed economici) 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.*

L'obiettivo deve essere raggiunto attraverso lo sfruttamento dei contenuti informativi presenti nel SIM, garantendo:

- la valorizzazione dei contenuti informativi di fonti amministrative di dati socio-economici anche gestite da enti regionali e locali;
- lo sviluppo di relazioni continue con le amministrazioni centrali, gli enti regionali e locali e le loro associazioni e più in generale con utenti interni ed esterni all'Istituto;
- la progettazione di procedure e strumenti idonei ad assicurare un'efficace fornitura dei dati agli utenti;
- la definizione di regole di comunicazione dei dati che assicurino il rispetto della normativa in materia di segreto statistico e protezione dei dati personali nell'ambito del Sistan e nei confronti di soggetti ad esso esterni, nonché la progettazione di un ambiente di diffusione dotato di metodi e strumenti in grado di assicurarne il rispetto.

Il progetto quindi opera un'**inversione** del flusso delle informazioni elementari. Di norma, i microdati dagli enti (visti come soggetti che acquisiscono e immagazzinano le dichiarazioni elementari) vengono trasmessi all'Istat che, una volta integrati/trattati/elaborati statisticamente, li rilascia in forma di informazione aggregata. Nel caso del progetto Archimede è l'Istat che potrà rilasciare collezioni di microdati la cui potenzialità informativa potrà essere sfruttata, a fini statistici, dagli stessi enti (che li potranno anche elaborare autonomamente) per le loro specifiche esigenze di informazione statistica. Lo stesso tipo di flusso potrà essere tecnicamente reso possibile verso qualificati centri di ricerca pubblici e privati per finalità di analisi sociali, economiche e demografiche e valutazioni indipendenti di politiche pubbliche a livello nazionale, regionale e locale.

### 1.3 Una modifica nella visione della produzione di informazioni statistiche

Nella sostanza il progetto Archimede opera uno *scouting* dei dati amministrativi disponibili per produrre collezioni statistiche di dati elementari da rilasciare ai soggetti del Sistan o a supporto della ricerca.

Considerando un “generico” processo di produzione statistica - in particolare utilizzando il modello GSBPM<sup>3</sup> - la prima fase identificata è la “*Specify needs*” la quale identifica un bisogno informativo statistico non soddisfatto e definisce come un sistema statistico nazionale possa soddi-

<sup>1</sup> A cura di G. Garofalo.

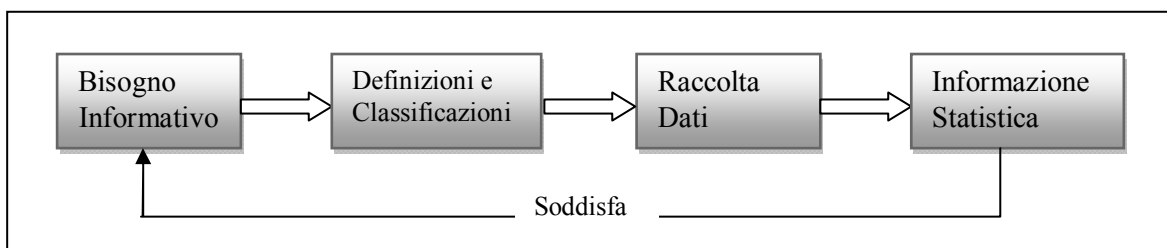
<sup>2</sup> La costituzione del Progetto come struttura organizzativa dell'Istituto è stata deliberata dal Consiglio dell'Istituto nella riunione del 14/01/2013 Cfr. Delibera 12/ DGEN del 18/01/2013.

<sup>3</sup> Altri modelli concettuali, quali il Information Systems Architecture for National and International Statistical Offices dell'UN del 1999 o il Data Life Cycle di Eurostat, pur con differenti livelli di approfondimento, propongono nella sostanza le stesse fasi.

sfarlo. Sostanzialmente è la fase che si conclude elaborando i “Business Case” dove si identificano gli obiettivi, i concetti rilevanti e le variabili, loro definizioni e classificazioni, per le quali i dati devono essere raccolti.

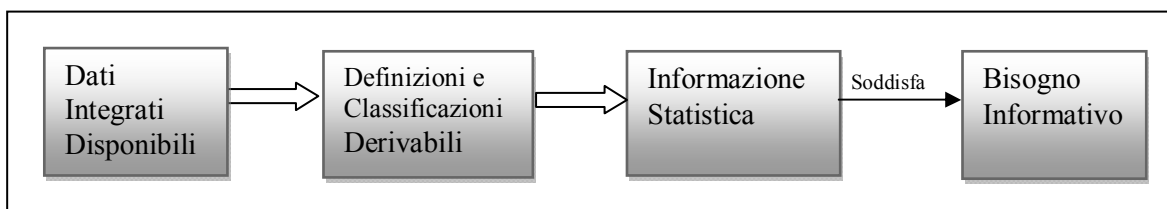
Da questa fase sono determinate le altre: il disegno delle attività, il processo di raccolta dei dati, i processi di elaborazione, integrazione e controllo, l’analisi, la diffusione, la valutazione etc.

Lo schema è sostanzialmente il seguente:



Questa struttura concettuale, che dai “bisogni” individua i “dati”, è invariante, che ci si riferisca ad un processo basato su indagini statistiche o su dati amministrativi<sup>4</sup> o su un processo misto.

Il progetto Archimede produce, concettualmente, una inversione di questo approccio. I “dati” **esistono**, e sono depositati in maniera integrata in SIM ed è l’esplorazione dell’informazione disponibile a identificare l’informazione statistica che può soddisfare un bisogno.



Questo approccio modifica il paradigma dell’uso delle fonti amministrative a fini statistici, in particolare in relazione a due aspetti non secondari del processo di produzione statistica. Il primo aspetto è che *le definizioni e le classificazioni possono essere determinate all’interno del processo di esplorazione e quindi non fissate a priori*. Il secondo, derivato dal primo, è che *l’adeguatezza dell’informazione prodotta rispetto agli schemi concettuali statistici può essere valutata solo ex-post*.

Il risultato più evidente di questo approccio è che, ad esempio, gli output prodotti da Archimede possono non essere strettamente vincolati ai sistemi di regolamentazione statistica europea o ai concetti del censimento.

Inoltre, a differenza dell’approccio “statico” della classica produzione statistica (dati confrontabili nel tempo e nello spazio), il progetto si caratterizza per la “dinamicità” dell’output prodotto, dal punto di vista definitorio e classificatorio. Infatti, l’inserimento nel SIM di nuove basi informative può modificare nel tempo il significato dell’informazione prodotta. Si pensi ad esempio al caso della definizione adottata di precarietà lavorativa (vedi il relativo report del progetto sperimentale): il valore informativo statistico del prodotto basato esclusivamente sulle tipologie contrattuali muta profondamente se si inseriscono anche informazioni sul reddito o sulle mono/pluricommitenza o sul livello di formazione del lavoratore.

L’assenza di vincoli *a priori* e la dinamicità, se da un lato permettono una *maggiore flessibilità informativa*, dall’altro rendono necessario *governarne la potenziale instabilità*.

La riflessione su quest’ultimo aspetto è ancora agli inizi, ma si evidenziano già due necessità. In primo luogo il sistema standard di “documentazione” dei prodotti deve essere il più completo e cor-

<sup>4</sup> Ovviamente, nel caso di statistiche totalmente o parzialmente derivate da fonti amministrative, per raccolta dei dati s’intendono anche tutte le attività di “traduzione” del dato amministrativo in dato statistico.

retto possibile, poiché deve guidare l'utente ad un corretto utilizzo dei prodotti stessi. In secondo luogo è necessario investigare tecniche, in particolar modo quelle classificatorie, non usualmente utilizzate dalla statistica pubblica (ad es. *Data Mining*) che possano ridurre – o mettere in evidenza – l'instabilità informativa.

#### 1.4 Le macrofunzioni del progetto

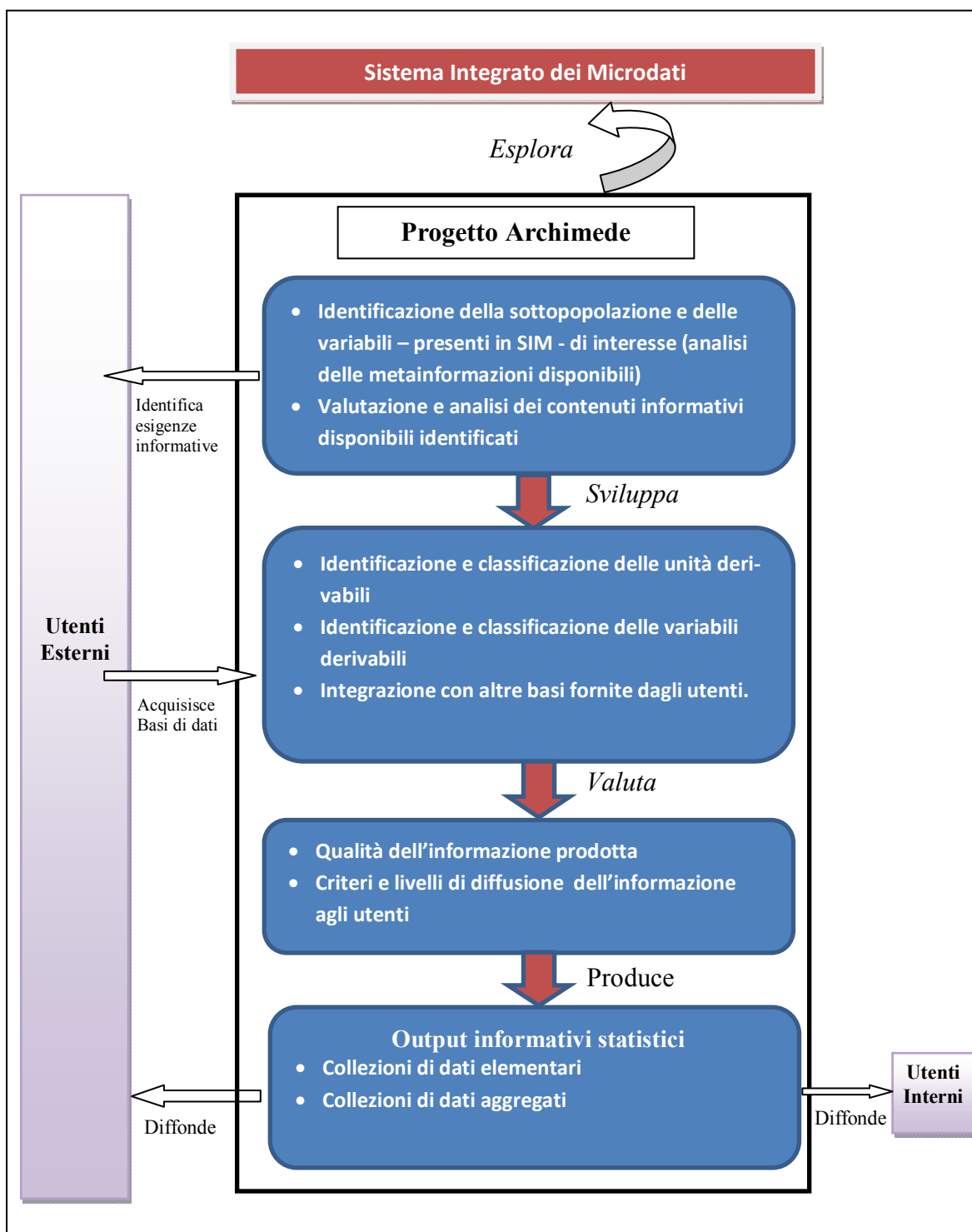
La Figura 1.1 rappresenta una schematizzazione delle macrofunzioni del progetto che possono essere sintetizzate nella:

1. identificazione delle esigenze informative di utenti interni ed esterni all'Istituto;
2. esplorazione delle informazioni contenute in SIM per soddisfare le esigenze;
3. realizzazione di output informativi – documentati e valutati in termini di qualità.

L'analisi delle esigenze si realizza attraverso lo sviluppo di relazioni continue con le amministrazioni centrali, gli enti regionali e locali e le loro associazioni e più in generale con utenti interni ed esterni all'Istituto. Ovviamente in questo senso un ruolo fondamentale è quello del SISTAN, come collettore di esigenze emergenti sulla base delle richieste che i singoli enti fanno al proprio Ufficio di Statistica. È evidente però che specifiche domande di informazione possono essere identificate dagli utenti anche sulla base dell'offerta che si rende disponibile. In tal senso è essenziale una informazione completa agli utenti sulla offerta di dati disponibile in SIM per tramite di ARCHIMEDE.

Definiti (o stimolati) i bisogni informativi degli utenti interni ed esterni, il progetto si propone di identificare e valutare le informazioni fruibili sulla base dei metadati messi a disposizione dal SIM e di una analisi dei contenuti effettivi. Al fine di soddisfare le specifiche esigenze conoscitive, può essere effettuato un trattamento delle fonti per derivare unità (combinazione o trasformazione delle unità presenti in SIM) variabili e loro classificazioni. L'eventuale integrazione di banche dati messe a disposizione dall'utente (ad esempio le prestazioni sociali erogate dai Comuni o gli utenti di corsi di formazione professionale regionali) può produrre nuovo valore aggiunto informativo per l'analisi territoriale.

Figura 1.1 - Le macrofunzioni del Progetto Archimede



## 1.5 Gli Output Informativi

Le tipologie di prodotti che il progetto deve realizzare sono tre:

- a) Collezioni di microdati integrati. Possono essere di due tipi:
1. *On demand*: realizzate in conformità alle specifiche esigenze informative che provengono da utenti esterni (amministrazioni, enti, centri di ricerca, ecc.).
  2. *Generalizzate*: realizzate come offerte informative progettate sulla base delle potenzialità del SIM, volte a soddisfare esigenze di un'ampia platea di utenti (anche interni all'Istituto) o di "gruppi omogenei" di utenti.

Le collezioni di microdati integrati hanno le seguenti caratteristiche:

- fanno riferimento a un insieme di unità che possono non rappresentare l'universo di una specifica popolazione;
  - utilizzano dati amministrativi di SIM *anche* non trattati statisticamente;
  - utilizzano concetti *non* necessariamente coerenti con la statistica ufficiale (es. Regolamenti europei).
- b) Collezioni di macrodati (tavole statistiche, indicatori, ...) derivabili dai microdati. Il passaggio da collezioni di microdati a macrodati risponde ad esigenze informative di utenti esterni che per varie ragioni – di riservatezza, organizzative, assenza di esperienze - non hanno la sufficiente autonomia per elaborare in proprio i dati elementari.
- c) Sottosistemi informativi statistici. Sono prodotti statistici che realizzano un incremento del patrimonio informativo dell'Istituto e del Sistema Statistico Nazionale. Essi devono avere le seguenti caratteristiche:
- fare riferimento ad una definita popolazione statistica (di individui, famiglie, unità economiche); utilizzare, per la diffusione dei dati, i caratteri di base definiti dalle popolazioni statistiche (età, sesso, luogo di nascita, residenza per gli individui; attività economica, dimensione, natura giuridica per le imprese); utilizzare in maniera integrata informazioni di dati amministrativi disponibili in SIM, trattati statisticamente (es. *editing and imputation*, modelli di stima probabilistici, ecc.); archiviare e diffondere i risultati tramite i canali ufficiali dell'Istituto (Armida verso l'interno, Adele verso l'esterno, i.Stat, ecc.) con la documentazione e le analisi di qualità standard dei prodotti statistici.

I prodotti sviluppati devono avere le seguenti caratteristiche:

- essere replicabili nel tempo e nello spazio (e quindi industrializzati);
- garantire output flessibili, in riferimento a differenti tipologie di utenti;
- permettere analisi specifiche per sottopopolazioni di interesse.

In particolare, avere a disposizione il SIM che tiene conto dell'evoluzione nel tempo delle basi dati amministrative, permette di sviluppare analisi di tipo longitudinale per le sottopopolazioni individuate (trasformazioni nella condizione lavorativa degli occupati, nella condizione sociale ed economica della famiglia, nei percorsi formativi, nella mobilità territoriale degli individui). Inoltre, la disponibilità di "universi" di popolazioni amministrative permette la diffusione di informazioni sia per ambiti territoriali sufficientemente piccoli (comunali e sub-comunali) sia per ambiti differenti da quelli amministrativi (Sistemi Locali del Lavoro).

Per garantire un corretto utilizzo dell'informazione diffusa da parte degli utenti, soprattutto in relazione alla programmazione e valutazione delle politiche, i prodotti (in particolare quelli individuati ai punti a e b) devono essere corredati da un sistema di documentazione e di valutazione sintetica della qualità.

## 1.6 La tutela della riservatezza nel trattamento e diffusione dei dati

Da un punto di vista generale Archimede rientra in quella tipologia di progetti statistici, coerenti con le vigenti disposizioni europee<sup>5</sup> e nazionali,<sup>6</sup> che favoriscono l'utilizzo e l'integrazione di dati provenienti da fonti amministrative rispetto alla conduzione di una nuova indagine statistica, in modo da ridurre tempi, costi e fastidio statistico sui rispondenti e da ampliare l'informazione statistica riferita a livelli territoriali di dettaglio altrimenti non indagabili se non a costi eccessivi.

Tuttavia, per le sue caratteristiche, il progetto deve assicurare il rispetto di specifiche garanzie nel trattamento e nella diffusione dei dati. A questo riguardo si sottolinea come i competenti uffici dell'Istat (Direzione Centrale per gli affari istituzionali, giuridici e legali – DCIG) siano stati coinvolti per lo studio dei profili giuridici e organizzativi del progetto.

In particolare le collezioni di dati elementari prodotte dal progetto Archimede avranno le caratteristiche vincolanti:

- di essere considerati dati statistici (in quanto frutto di processi di integrazione, classificazione, stima che sfruttano metodologie statistiche) e quindi diffondibili esclusivamente a soggetti SISTAN;
- di non contenere caratteri identificativi (Codice Fiscale o nome e cognome), ma un codice identificativo (numerico progressivo) assegnato dal SIM;
- Nondimeno si evidenzia che esse:
  - faranno riferimento ad intere sottopopolazioni (e non quindi a campioni statistici di popolazioni);
  - potranno contenere dati sensibili (ad es. persone con disabilità);
  - permetteranno di ottenere dai dati amministrativi integrati variabili derivate, non tratte da informazioni “dichiarate” dall'individuo o unità economica all'ente che direttamente raccoglie per finalità amministrative l'informazione elementare;
- anche in assenza di caratteri identificativi potrebbero permettere l'identificazione degli individui in ambiti territoriali ristretti (piccoli comuni) a causa della molteplicità d'informazioni disponibili (studio, lavoro, tipologia familiare, ecc.) per ciascun soggetto.

Pertanto, in relazione alle peculiarità del progetto sarà necessario definire *ex-ante* le regole per il trattamento e la diffusione di dati individuali che garantiscano il rispetto delle norme poste a tutela dei dati personali e del segreto statistico. Poiché nelle collezioni di microdati integrati ci possono essere informazioni che potrebbero essere utilizzate da parte delle amministrazioni del Sistan per finalità di controllo e vigilanza (*intelligence*) amministrativa, sarà necessario approntare adeguati mezzi di vigilanza sull'uso che ne faranno gli uffici di statistica richiedenti.

## 1.7 La progettazione dei sistemi di diffusione dei dati

La diffusione dei prodotti del progetto non può prescindere dalla utilizzazione, nel breve periodo, dei sistemi attualmente sviluppati dall'Istituto: ARMIDA, Laboratorio ADELE, I.STAT. I primi due per la diffusione delle collezioni di microdati il terzo per dati aggregati.

Alcuni aspetti oggettivi rendono, però, necessaria la valutazione, da parte dell'Istituto, dell'opportunità di realizzare un diverso sistema di diffusione. Tali aspetti fanno riferimento allo specifico contenuto informativo dei prodotti, alla necessità di operare integrazioni con basi di dati esterne all'Istituto, alle specifiche problematiche di riservatezza evidenziate nel paragrafo precedente.

La figura 1.2 schematizza, in maniera generale, i contenuti connessi alla diffusione delle informazioni prodotte dal progetto.

<sup>5</sup> In particolare le proposte di modifica della Legge statistica europea – Reg (CE) 223/2009 – introducono nuove disposizioni in materia di accesso ai dati amministrativi, di utilizzo e integrazione degli stessi, con il principale scopo di creare un quadro giuridico che permetta un utilizzo più ampio delle fonti amministrative per la produzione delle statistiche europee, riducendo gli oneri per i rispondenti, gli Istituti di Statistica e le altre autorità Nazionali.

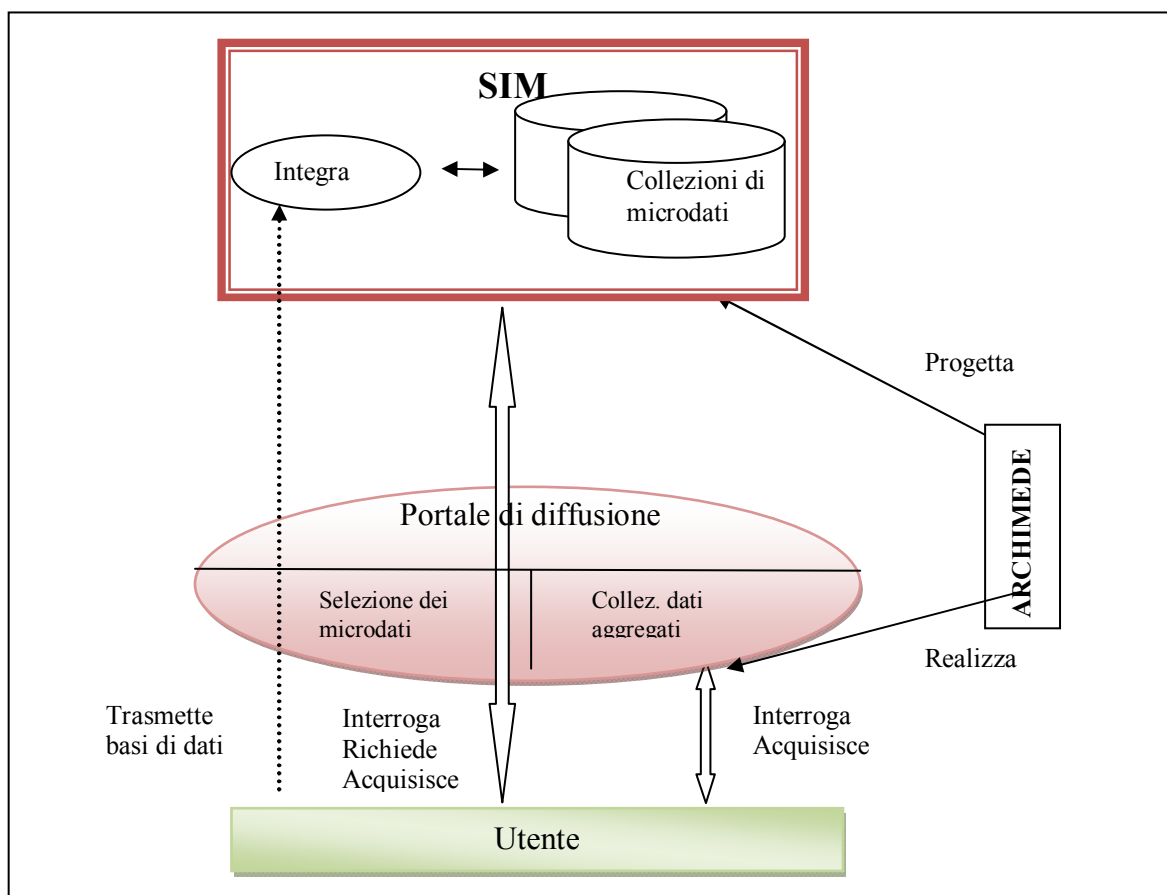
<sup>6</sup> D.lgs 322/1989, Legge n. 681/1996.

Un primo elemento è quindi connesso alla specificità dei prodotti del progetto Archimede. Essi, infatti, sono il risultato di una esclusiva integrazione di fonti amministrative, da cui possono derivare sistemi di classificazione e definizione non “omogenei” con quelli della statistica ufficiale – nazionale ed europea – e per i quali la qualità è più connessa ad un concetto di “servizio” a specifiche esigenze informative. In qualche modo quindi essi possono essere considerati come “diversi” da quelli della statistica ufficiale. E’ necessario quindi valutare se la loro diffusione, sia a livello di microdati sia a livello di dati aggregati, non debba essere anche fisicamente separata per connotare meglio la loro specifica caratteristica.

Il secondo aspetto risiede nella necessità, connessa al concetto di “servizio” informativo, di integrare specifiche banche di dati disponibili presso gli enti, in primo luogo Comuni e Regioni, che vogliano utilizzare i prodotti di Archimede ad integrazione delle informazioni in loro possesso. Si pensi ad esempio alla integrazione di alcune prestazioni, di natura sociale e assistenziale a sostegno delle famiglie, che possono rientrare nei redditi da trasferimenti pubblici, non disponibili per l’Istituto. Un sistema di diffusione deve prevedere anche lo sviluppo di funzionalità che permettano questo tipo di integrazione garantendo, anche, tutti gli aspetti connessi al segreto statistico (in particolare la non identificabilità dei soggetti per i quali le differenti informazioni si integrano, accennati nel paragrafo precedente).

Infine, un sistema che debba garantire congiuntamente sia una diffusa utilizzabilità dell’informazione sia la massima garanzia di riservatezza, non può prescindere dallo sviluppo delle opportune funzionalità di controllo (identificazione, riconoscimento, autorizzazione, archiviazione) dei soggetti che richiedono il servizio e delle informazioni richieste. Per garantire maggiormente il rispetto del segreto statistico, tali funzionalità dovrebbero, anche, prevedere differenti livelli di autorizzazione alla acquisizione di dati elementari sulla base delle caratteristiche dell’ente richiedente e alle motivazioni della richiesta.

Figura 1.2 - Il sistema di diffusione





## 1.8 I sotto-progetti sperimentali realizzati

Al fine di permettere una valutazione complessiva del progetto Archimede, sia rispetto alla sua capacità di “*ampliare l’offerta informativa dell’Istat*” (sfruttando in maniera esclusiva dati amministrativi integrati), sia rispetto alle problematiche connesse all’organizzazione, alle interrelazioni con altre strutture dell’Istat e con Enti esterni, alle risorse e ai costi, è stato deciso di sviluppare alcuni progetti sperimentali.

In generale, il processo di produzione messo in atto si è svolto in una fase di riorganizzazione del DICA, in conseguenza della quale il gruppo di lavoro impegnato nel progetto ha dovuto farsi carico di attività, quali l’integrazione fisica delle fonti, la cui responsabilità (e competenza) spetta ad altre strutture del Dipartimento. Per questo motivo, la valutazione dei risultati deve tener conto del fatto che essi sono stati ottenuti attraverso processi non industrializzati.

I tre sotto-progetti sperimentali sono stati individuati sia nel corso di incontri (informali) avuti con utenti “qualificati”,<sup>7</sup> nei quali sono emersi alcuni bisogni informativi, sia attraverso una prima ricognizione delle fonti amministrative già disponibili in Istituto, o di recente acquisizione.

*Popolazione insistente*: l’identificazione e la quantificazione della popolazione che usa il territorio è fra i temi più sentiti dalle amministrazioni locali. Infatti, è questa la popolazione che determina squilibri finanziari e logistici per i governi locali nell’ambito territoriale che essi amministrano. L’obiettivo principale del progetto è identificare e quindi quantificare le componenti della popolazione caratterizzate da diversa residenzialità: i residenti “effettivamente” dimoranti, i temporaneamente dimoranti e i pendolari giornalieri. L’obiettivo è perseguito specificando la relazione di stato esistente tra individui e unità territoriali (luogo di lavoro in X, di studio in Y, di residenza in Z, ecc.) e la relazione di mobilità verso questi luoghi.

Il progetto è basato sull’identificazione di un sistema informativo del tipo *Persons&Places* che collega a ciascun individuo una stringa di informazioni territoriali che ne identificano la residenza (da anagrafi comunali), il domicilio fiscale (da anagrafi fiscali), il luogo di lavoro (da anagrafi previdenziali e assicurative), il luogo di studio (da anagrafi del Miur). Dal sistema è possibile derivare, per un determinato territorio e per sottopopolazioni, matrici di Origine/Destinazione da cui quantificare la popolazione insistente e definire specifici indicatori (ad es. indicatori di attrazione e di autocontenimento).

La sperimentazione è stata effettuata per l’intero territorio nazionale.

*Precarietà lavorativa*: se la principale problematica connessa al mercato del lavoro di un Paese è ben descritta attraverso lo studio della disoccupazione, la presenza di posti di lavoro non stabili – la loro misura e descrizione – rappresenta oggi in Italia una tematica al centro di numerose analisi, sia economiche che sociali. L’obiettivo principale del sotto-progetto è di effettuare l’analisi di un universo di individui definiti “lavoratori precari”, registrandone aspetti socio-anagrafici e caratteri economico-lavorativi in un arco temporale e di seguirne le trasformazioni nel tempo. Per lavoratore precario s’intende un individuo che, in un determinato periodo, svolge una o più categorie di lavoro definite precarie, anche contemporaneamente, e non svolge alcuna categoria di lavoro stabile. Il lavoro precario è definito come un’attività lavorativa caratterizzata da contratti di (1) lavoro dipendente a termine, (2) lavoro assimilato a dipendente, (3) lavoro autonomo nei casi che presentano caratteristiche di subordinazione impropria.

La base di riferimento del progetto è stata costituita dal registro ASIA-Occupazione, realizzato dall’Istat per il Censimento 2011, che lega ciascun lavoratore a ciascuna impresa per mezzo di un determinato contratto di lavoro. In questa base informativa sono state integrate le informazioni degli assicurati ex-Inpdap (in gran parte pubblico impiego), dei lavoratori domestici e dei lavoratori autonomi dell’agricoltura. La sperimentazione si è concentrata sulla costruzione di informazioni micro in serie storica che permettono un’analisi longitudinale sulle transizioni (entrate/uscite) della condizione lavorativa “precaria” in “stabile” e viceversa.

<sup>7</sup> I primi incontri sono stati effettuati con alcuni responsabili degli uffici Regionali e Comunali di Statistica a Milano, Firenze e Bologna.

*Condizioni socio-economica delle famiglie:* L'interesse al tema deriva dall'opportunità di costruire una struttura informativa sulle famiglie che consenta di analizzare vari aspetti della loro condizione socio-economica. Le esigenze conoscitive riscontrate presso testimoni privilegiati fanno riferimento, in particolare, alla declinazione del concetto di precarietà socio-economica e quindi alla sua scomposizione in elementi osservabili e misurabili. L'obiettivo principale del sotto-progetto è classificare tutte le famiglie residenti in base a caratteristiche statisticamente osservabili, che descrivano le loro eventuali condizioni di precarietà sociale ed economica. In particolare, la condizione socio-economica è stata analizzata attraverso cinque dimensioni: tipologia familiare, reddito, condizione lavorativa, disagio sociale, istruzione e formazione. È stata progettata e realizzata una ricostruzione del reddito individuale (e familiare) sulla base dell'integrazione di fonti fiscali e previdenziali.

In modo sintetico il prospetto successivo descrive le attività svolte nel corso del 2013.

### Prospetto 1.1 - Sintesi delle attività svolte nel corso del 2013

Periodo	Attività	Azioni	Risultati
Marzo/Aprile	Posizionamento del progetto: obiettivi e funzioni del Progetto, relazioni funzionali con altre strutture del Dipartimento Identificazione dei sottoprogetti sperimentali	Incontri informali con "utenti" esperti. Prima analisi delle fonti disponibili Incontri con strutture DICA (ABA/REG...) Analisi del GSBPM	
Maggio	Specificazione e articolazione dei sottoprogetti	Identificazione nuove fonti non ancora o parzialmente acquisite Approfondimenti contenuti informativi fonti disponibili	<i>Realizzazione dei documenti di "Business case"</i> Presentazione e discussione dei sottoprogetti ad alcuni responsabili di Uffici di Statistica comunali e regionali (Firenze)
Giugno/Luglio	Sviluppo del disegno dei sottoprogetti (identificazione completa delle fonti/variabili, identificazione del processo, approfondimenti metodologici)	Incontri informali con DISA e sue strutture (EU-SILC, Pensioni, Formazione, Disabilità) Incontro con CNR - Un. di Pisa sui BIG DATA Incontro Reg. Lombardia /CRISP su redditi/mercato del lavoro	<i>Realizzazione dei documenti di "Design"</i> Presentazione e discussione del progetto Archimede al Seminario "interno" del DICA
Agosto	Inizio acquisizione delle basi di dati per lo sviluppo dei sottoprogetti.		
Settembre/ Novembre	Completamento acquisizione dati Sviluppo processo dei sottoprogetti sperimentali Prime valutazioni dei risultati	Realizzazione di un documento di progetto per la sperimentazione dei BIG DATA per il progetto popolazione insistente. Costituzione di un gruppo di lavoro informale del DICA (Iac/Ada/Archimede) per lo sviluppo di applicazioni di Data Mining per il sottoprogetto vulnerabilità familiare Incontro con esperti dell'ind. Forze di Lavoro sui primi risultati del sottoprogetto precarietà lavorativa	Presentazione del progetto Archimede al Convegno USCI (Messina) Presentazione e discussione della metodologia e dei primi risultati sulla ricostruzione del reddito familiare ad alcuni responsabili di Uffici di Statistica comunali (Bologna) Presentazione dei primi risultati della sperimentazione "precarietà lavorativa" al workshop ISTAT/Min. Lavoro (Roma)
Dicembre	Completamento della sperimentazione	Primi incontri con la DCIG sulle problematiche della <i>privacy</i> nella realizzazione e diffusione dei prodotti di Archimede Incontro con IAC/ADA per la progettazione dei sistemi di diffusione	La commissione BIG DATA inserisce sottoprogetto <i>persons &amp; places</i> tra quelli di suo interesse primario Presentazione del Progetto Archimede al Convegno URBES (Roma) <i>Realizzazione della prima versione del documento: Progetto Archimede - i risultati sperimentali e le linee di sviluppo</i>

## 2. Il sotto-progetto sperimentale: popolazione insistente<sup>8</sup>

### 2.1 Le caratteristiche della sperimentazione

#### 2.1.1 Obiettivi

La popolazione che insiste su un territorio si distingue dalla popolazione residente tanto più è forte la capacità del territorio di attrarre persone residenti altrove, e tanto più debole è quella di contenere le persone residenti.

A livello nazionale non esistono ancora dati ufficiali, e il progetto nasce proprio per studiare il grado di fattibilità del recupero di questo gap informativo, congiuntamente alla sostituzione totale o parziale della tradizionale rilevazione censuaria sugli spostamenti quotidiani con dati amministrativi.

Questi obiettivi possono essere meglio declinati nel seguente modo:

Ob.1 - Determinare la matrice Origine/Destinazione (OD), a partire da un sistema informativo, denominato Persons&Places (P&P), sulla popolazione che per motivi di studio o di lavoro si sposta dal comune di residenza.

Ob.2 – Qualificare, attraverso la stima della frequenza degli spostamenti, il tipo di mobilità come “giornaliera” o “periodica”, e classificare l’individuo nelle diverse modalità di *city user*.

Ob.3 - Quantificare la popolazione insistente.

#### 2.1.2 Unità, riferimento temporale, e principali definizioni adottate nel prototipo del sistema informativo Persons&Places\_2011

Il progetto è incardinato sulla disponibilità di molteplici fonti di dati amministrativi sul lavoro e sullo studio (archivi che non sono ancora entrati nelle analisi sulle direttrici della mobilità e sulle aree di gravitazione). Premessa dello sfruttamento è che essi siano prima integrati tra loro in un sistema pertinente e non ridondante, composto di unità identificate univocamente.

Il **Sistema Informativo** in cui le fonti amministrative individuate sono state integrate, è denominato *Persons & Places*.

L’**universo di riferimento** è dato dall’unione delle persone residenti in Italia, delle persone che lavorano in Italia, e delle persone che studiano in Italia. Nel sistema quindi coesistono diverse popolazioni derivate da fonti “indipendenti” tra loro: la popolazione residente, quella degli studenti e degli universitari, quella dei lavoratori.

Le **Unità statistiche** dell’Obiettivo n.1 sono:

**Persona Residente:** persona iscritta in anagrafe, così come risulta dalle Liste Anagrafiche Comunali - LAC al 9 ottobre 2011

**Lavoratore:** persona presente in almeno uno tra gli archivi ASIA - occupazione 2011, ex-INPDAP, Lavoratori Domestici, Lavoratori Autonomi dell’Agricoltura, Anagrafe personale docente e non docente delle università.

**Studente:** persona presente in Anagrafe degli studenti non universitari o Anagrafe degli studenti universitari.

Le Unità di analisi dell’Ob.2 e Ob.3 sono:

**SR = Standing resident, (Persona Stanziale nel comune C<sub>i</sub>),** persona che ha una attività di lavoro/studio nel comune C<sub>i</sub> di iscrizione anagrafica (SR<sup>att</sup>(i)); o persona iscritta nell’anagrafe C<sub>i</sub> che non ha attività di studio/lavoro (SR<sup>natt</sup>(i)).

**EC = Embedded city user, (Temporaneamente Dimorante in C<sub>i</sub>),** persona che lavora o studia nel Comune C<sub>i</sub> - in cui non ha residenza anagrafica – e vi alloggia;

**DC = Daily city user, (Pendolare Quotidiano verso C<sub>i</sub>),** persona che lavora o studia nel Comune C<sub>i</sub> - in cui non ha residenza anagrafica - recandovisi giornalmente;

<sup>8</sup> I §§ 2.1, 2.4 e 2.5 sono a cura di R. Vivio, i §§ 2.2 e 2.3 sono a cura di P. Barberis, l’appendice è a cura di M.D. De Gabrieli.

**FC = Free city user (Persona di Contatto con C<sub>i</sub>)**, persona che occasionalmente si è recato in C<sub>i</sub> - in cui non ha residenza anagrafica - per lavoro, incontri d'affari, meeting, fiere, esposizioni, turismo, ecc.

Equazione della Popolazione residente in Anagrafe (RA) del Comune C<sub>i</sub> in funzione delle unità di analisi dell'obiettivo 2 e 3:

$$RA_{(i)} = \sum SR_{(i)}^{att} + \sum SR_{(i)}^{natt} + \sum RA_{(i),EC_{(j)}} + \sum RA_{(i),DC_{(j)}}$$

dove:

$SR_{(i)}^{att}$  = Residente che lavora/studia nello stesso comune C<sub>i</sub> di residenza anagrafica;

$SR_{(i)}^{natt}$  = Residente che non ha attività di studio/lavoro

$RA_{(i),EC_{(j)}}$  = Residente anagrafico in C<sub>i</sub>, temporaneamente dimorante in C<sub>j</sub> per attività di studio/lavoro

$RA_{(i),DC_{(j)}}$  = Residente anagrafico in C<sub>i</sub>, pendolare giornaliero verso C<sub>j</sub> per attività di studio/lavoro.

Alla formazione della **Popolazione che insiste sul territorio di un comune** (Ob.3) concorrono gli *Standing resident*, i *Daily users* e gli *Embedded city users*: ma mentre gli *standing* insistono su un solo comune, i *daily* e gli *embedded* insistono sia sul comune Origine, che sul comune Destinazione sede della loro attività. Quindi la misurazione della Popolazione Insistente non può che passare per il tempo e la durata della permanenza in entrambi i luoghi Origine e Destinazione.

Come approccio iniziale per il test del 2011 si è scelto quello di contare le “teste” nel comune di Destinazione, sottraendole da quello di Origine:

Popolazione insistente nel Comune in C<sub>i</sub>:

$$PI_{(i)} = \sum SR_{(i)} + \sum RA_{(i),EC_{(i)}} + \sum RA_{(i),DC_{(i)}} - [\sum RA_{(i),EC_{(j)}} + \sum RA_{(i),DC_{(j)}}]$$

Nel corso del 2014 si affineranno i metodi per arrivare a poter calcolare la PI come somma ponderata in funzione anche della fascia oraria di permanenza sul territorio (diurna, notturna), e del tipo di giorno (feriale, festivo) di ciascuna persona:

$$PI_{(i)} = \sum SR_{(i)} + \alpha_{ji} \sum RA_{(j),EC_{(i)}} + \beta_{ji} \sum RA_{(j),DC_{(i)}} - [\alpha_{ij} \sum RA_{(i),EC_{(j)}} + \beta_{ij} \sum RA_{(i),DC_{(j)}}]$$

Dove  $\alpha$  e  $\beta$  sono i parametri che individuano la “quantità” di sfruttamento di un territorio. La popolazione insistente è il risultato del bilancio di mobilità pesata in un intervallo di tempo  $\Delta t$ .

Di seguito nella Matrice Origine / Destinazione sono evidenziate con diversi colori le combinazioni della coppia di comuni - che costituiscono l'origine di un flusso e la sua destinazione - a seconda del fatto che sia ammissibile o non ammissibile per una data tipologia di city user.

Per esempio:

- un individuo non può essere *Standing resident* in più di un comune contemporaneamente (combinazione non ammissibile);
- un individuo *Standing resident* non può essere *Embedded* o *Daily* in altri comuni (combinazione non ammissibile).

Come origine e come destinazione sono ipotizzati i generici comuni i-1, i, i+1 (e l'estero). Per ogni comune vengono considerati i tre diversi tipi di utilizzatori del territorio nei quali può essere distinta la popolazione residente in anagrafe (*Standing*, *Embedded* e *Daily*).

**Schema 2.1 - Matrice Origine / Destinazione per tipo di city user (Resident, Embedded e Daily) e comune**

COMUNI	Tipologie User	...	Comune i-1			Comune i			Comune i+1			...	ESTERO			TOTALI
		...	SR	EC	DC	SR	EC	DC	SR	EC	DC	...	SR	EC	DC	
.....	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...
Comune i-1	SR	...	...	...	...	...	...	...	...	...	...	...	...	...	...	$\Sigma SR(i-1)$
	EC	...	...	...	...	...	...	...	...	...	...	...	...	...	...	$\Sigma EC(i-1)$
	DC	...	...	...	...	...	...	...	...	...	...	...	...	...	...	$\Sigma DC(i-1)$
Comune i	SR	...	...	...	...	...	...	...	...	...	...	...	...	...	...	$\Sigma SR(i)$
	EC	...	...	...	...	...	...	...	...	...	...	...	...	...	...	$\Sigma EC(i)$
	DC	...	...	...	...	...	...	...	...	...	...	...	...	...	...	$\Sigma DC(i)$
Comune i+1	SR	...	...	...	...	...	...	...	...	...	...	...	...	...	...	$\Sigma SR(i+1)$
	EC	...	...	...	...	...	...	...	...	...	...	...	...	...	...	$\Sigma EC(i+1)$
	DC	...	...	...	...	...	...	...	...	...	...	...	...	...	...	$\Sigma DC(i+1)$
⋮	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...
ESTERO	SR (AIRE)	...	...	...	...	...	...	...	...	...	...	...	...	...	...	$\Sigma SR(e)$
	EC	...	...	...	...	...	...	...	...	...	...	...	...	...	...	$\Sigma EC(e)$
	DC	...	...	...	...	...	...	...	...	...	...	...	...	...	...	$\Sigma DC(e)$

Legenda.

RC = Standing Resident o Resident city user,

EC = Embedded City user,

DC = Daily City User.

Combinazione ammissibile = casella bianca

Combinazione non ammissibile = casella rossa

### 2.1.3 Le fonti dei dati

Le fonti amministrative più funzionali al perseguimento degli obiettivi (popolazioni, unità statistiche e variabili), sono fonti acquisite dall'Istat nell'ambito dei vari Progetti istituzionali regolamentati all'interno del Piano Statistico Nazionale.

Le fonti sono state scelte seguendo come criterio quello della pertinenza e non eccedenza informativa rispetto alle finalità statistiche del sistema informativo. Sono integrabili tra loro, coerenti, accurate e tutte a copertura nazionale. Per la sperimentazione sono tutte riferite al 2011.

La caratteristica di lavoratore, studente o residente è verificata rispetto ai relativi archivi ufficiali riportati nel prospetto che segue.

**Prospetto 2.1 - Unità e localizzazione rispetto ai relativi archivi ufficiali**

Archivio / Registro	Unità e localizzazione
POPOLAZIONE_RESIDENTE Popolazione al censimento 2011 Popolazione presente sul territorio come lavoratore o studente e non censita	Residenti, e comune di residenza
LAVORATORI ASIA_Occupazione Posizioni assicurative ex-INPDAP Lavoratori Domestici Lavoratori Autonomi dell'Agricoltura Anagrafe del personale docente e non docente delle università	Lavoratori (residenti e non residenti), e comune di lavoro
STUDENTI e UNIVERSITARI Anagrafe degli studenti non universitari Anagrafe degli studenti universitari	Studenti (residenti e non residenti), e comune sede dell'istituto Universitari (residenti e non residenti), e comune sede del corso universitario

## 2.2 Il Sistema P&P

### 2.2.1 La struttura fisica

La struttura finale del sistema “*Persons&Places*” è formata da una matrice in cui le unità sono gli individui e le colonne principali sono i Comuni di residenza, lavoro e studio. Inoltre il sistema contiene informazioni sul comune di nascita e quello di domicilio fiscale di ciascun individuo.

Il sistema è costruito in modo tale da poter essere facilmente integrato con tutti gli archivi del SIM. Questa struttura *plug&play* si rivela strategica in tutte le situazioni in cui occorra fare degli approfondimenti mirati, perché li rende possibili con una semplice integrazione di dati da SIM. E’ questo, per esempio, il caso dello sviluppo di un modello di stima della probabilità di spostamento per un individuo che lavori (o studi) in un luogo diverso da quello di residenza: in questo caso l’aggancio con altre informazioni quali la dimensione demografica del comune, il codice ATECO dell’azienda in cui si lavora, le informazioni sulla famiglia, ecc. possono essere usate a fini esplicativi.

### 2.2.2 Le Fasi del processo

Per la costruzione del sistema informativo si è proceduto per fasi. Il processo è avvenuto parallelamente alle attività di implementazione del SIM, per ovvi motivi di contenimento dei tempi di realizzazione del prototipo (SIM a marzo 2013 era ancora in fase di costruzione), dunque si è cercato di mantenerne i principi ispiratori per ciò che concerne:

- l’identificazione di ciascun individuo/entità presente in più fonti mediante un “**Codice identificativo**” unico;
- la definizione per ciascuna **entità** delle relazioni logiche e fisiche (**nel tempo e nello spazio**) che intercorrono tra le informazioni ad essa riferite nelle diverse fonti;
- procedere all’integrazione avendo come pilastri i codici e le relazioni.

Il Codice Fiscale (principale variabile identificativa dell’individuo nei vari archivi) è stato verificato rispetto alla fonte che contiene i codici fiscali ufficiali, cioè la Base Dati Anagrafica dell’Anagrafe Tributaria del MEF.

La caratteristica di lavoratore, studente o residente è derivata in funzione della presenza dell’unità nell’archivio di riferimento.

La relazione dell’unità con i luoghi (comune di studio/lavoro/residenza) è derivata dagli archivi scelti.

Le fasi sono state:

#### **FASE 1. La popolazione dei residenti**

La popolazione di riferimento per i residenti nel sistema informativo a regime sarà quella delle persone iscritte in anagrafe integrate con tutti gli individui che, non residenti, lavorano o studiano in Italia. Per la versione prototipale, riferita all’anno 2011, si fa riferimento alla Popolazione Legale censita integrata con gli individui non censiti ma che lavorano o studiano in Italia (questi records sono determinati da regole di editing a fine processo; in questa fase vengono acquisiti tutti i records di persone non censite).

Le sottofasi sono state due:

- a) Acquisizione dei “censiti già iscritti come residenti in LAC” e dei “Censiti non iscritti come residenti in LAC”;
- b) Unione tra “risultato di fase a” e “Non censiti iscritti come residenti in LAC”.

#### **FASE 2. Codici fiscali mancanti, e codici fiscali non formalmente corretti nei microdati dell’archivio dei residenti**

a) Recupero dei Codici fiscali mancanti in LAC: i codici fiscali mancanti dei residenti in Lista Anagrafica Comunale (LAC) sono stati imputati utilizzando le informazioni contenute nell’Anagrafe Tributaria (AT). Questa situazione ha riguardato 905 mila unità, di cui 682 mila tra i nuovi censiti. In particolare è stata fatta la scelta di utilizzare, per i codici fiscali mancanti in LAC,

i codici fiscali presenti in AT nei casi di coincidenza esatta nei due archivi dei dati corrispondenti a nome, cognome e data (giorno, mese, anno) di nascita

Sono stati recuperati 401 mila codici fiscali.

b) “Quarantena” dei Codici fiscali non formalmente corretti: nel processo di costruzione del prototipo sono stati realizzati soltanto linkage deterministici, considerando come chiave di aggancio il Codice fiscale, e come criterio quello dell’uguaglianza esatta della chiave nei due archivi. Questa scelta è stata dettata dall’opportunità di privilegiare l’obiettivo del completamento del Sistema informativo per effettuare i test preventivati di utilità del sistema stesso, piuttosto che l’obiettivo della minimizzazione dei residui. Pertanto i codici non linkati perché non strettamente identici sono stati accantonati: nel processo a regime saranno sottoposti a un processo di linkage più raffinato per minimizzarne la numerosità.

**FASE 3. Unione della popolazione degli Studenti e del comune di studio ai microdati di Fase2:** l’unione di queste due popolazioni è un’operazione di linkage il cui risultato principale è quello di associare il comune sede della scuola agli studenti già presenti tra i microdati di fase 2, e, per gli studenti non già presenti, anche il record delle caratteristiche individuali. Alla fine della Fase 3 vengono accantonati i record associati a codici fiscali non identici, o mancanti, o multipli.

**FASE 4. Unione della popolazione degli Universitari e del comune di studio ai microdati di Fase 3:** idem a Fase 3. Come comune di studio - dato l’obiettivo dell’analisi della mobilità - è stato scelto quello della sede del corso piuttosto che quello della sede dell’ateneo, essendo più aderente al concetto di polo di attrazione dello spostamento fisico di uno studente. Per la localizzazione degli studenti si è fatto ricorso agli archivi del MIUR. Sia per gli studenti delle scuole sia per quelli universitari, la localizzazione ha richiesto un linkage preliminare tra due archivi del MIUR (quello dedicato alle strutture scolastiche con quello dedicato agli studenti; idem per l’università). Il comune di studio è definito per il 95% del totale degli studenti e degli universitari. Quindi - per il momento - nel sistema “Persons&Places” sono presenti il 95% delle persone che studiano in Italia, sia residenti che non residenti, e iscritte ad ogni ordine di scuola, dalle materne alle università.

#### **FASE 5. Identificazione della localizzazione dei lavoratori.**

Per l’identificazione del luogo di lavoro, si è ri-usato in primo luogo un sistema informativo dell’Istat: ASIA-Occupazione:<sup>9</sup> si tratta di una base di microdati integrati per impresa e occupato, costruita ai fini della realizzazione del CIS (Censimento Industria e Servizi 2011), nonché dell’aggiornamento del Registro Istat ASIA - Imprese attive.<sup>10</sup> La ragione di questa scelta è dettata dal fatto che in **il sistema contiene il luogo di lavoro a livello di unità locali della quasi totalità degli addetti** delle imprese dei settori dell’industria e dei servizi (oltre il 65% del totale degli occupati in Italia). Per i lavoratori non compresi in ASIA-Occupazione o per i quali non è stata identificata l’unità locale in cui operano, la localizzazione del luogo di lavoro è stata effettuata sperimentalmente dal Progetto Archimede. In particolare si tratta dei:<sup>11</sup>

- lavoratori presenti in fonti non integrate in ASIA-Occupazione ovvero assicurati presso EX INPDAP, Autonomi dell’Agricoltura e Lavoratori Domestici;

<sup>9</sup> È costituito dall’integrazione complessa di un insieme di fonti amministrative, dove l’unità di analisi (unità economica impresa) è identificata sulla base di definizioni condivise anche a livello europeo, e l’occupazione è considerata un attributo di tale unità. La struttura informativa si compone di tre livelli informativi (impresa, lavoratori e relazioni tra lavoratori e imprese) l’attività lavorativa viene svolta. Vedi anche: Istat, Servizio REG (2013).

<sup>10</sup> Archivio statistico delle imprese attive: rappresenta le imprese che hanno svolto un’attività produttiva nell’anno ed è il risultato dell’integrazione di numerose fonti amministrative. La produzione degli archivi riguarda le unità economiche attive (imprese, UL) e i relativi caratteri anagrafici (denominazione) e di struttura (attività economica, dimensione, forma giuridica), nonché di localizzazione e demografici.

<sup>11</sup> Per costoro la localizzazione è stata identificata per il mese di Dicembre 2011, così da garantire l’allineamento temporale con i dati derivati da ASIA-Occupazione.

- dipendenti occupati in istituzioni non profit;
- tutti i parasubordinati e lavoratori interinali.

Il processo realizzato per identificare il luogo di lavoro dei suddetti lavoratori viene descritto nell'approfondimento di paragrafo 2.5. Dalla tavola 2.1 si evidenzia come il comune di lavoro è stato identificato per circa il 59% di tali lavoratori. Tale risultato è in gran parte dovuto allo scarso successo ottenuto nel caso dei collaboratori e lavoratori interinali.

**Tavola 2.1 - Lavoratori non compresi nel Sistema Informativo di ASIA-Occupazione (\*), per tipologia lavorativa**

TIPOLOGIA	Lavoratori		
	Totale	con contratto attivo a dicembre e localizzati	
		Valore assoluto	Percentuale sul totale
Dipendenti	5.730.104	4.033.751	70,40
Interinali	479.938	20.147	4,20
Parasubordinati	1.024.297	232.828	22,73
<b>Totale</b>	<b>7.234.339</b>	<b>4.286.726</b>	<b>59,26</b>

(\*) Persone occupate nella Pubblica amministrazione, occupate in settori non censiti dal Censimento dell'Industria e dei servizi, occupate come lavoratori domestici, occupate come autonomi in agricoltura.

In conclusione, il sistema “*Persons&Places*” contiene per la prima volta insieme circa l’85% delle persone che lavorano in Italia, sia residenti che non residenti, sia pubblici che privati. In P&P è contenuto un Registro nazionale dei Lavoratori insistenti.

#### FASE 6. Editing finale

Il processo non ha previsto una vera e propria stesura di un piano di *editing*, ma soltanto l'applicazione di alcune regole di base, tra cui le seguenti:

- **transcodifica delle variabili ed omogeneizzazione:** ha riguardato soprattutto le variabili territoriali. Spesso negli archivi amministrativi è utilizzato il codice catastale piuttosto che quello Istat;
- **trattamento dei deceduti:** sono state identificate (ed eliminate) le persone che secondo la base dati anagrafica del MEF erano decedute prima della data di riferimento del sistema, ma non ancora cancellate dagli archivi assemblati per la costruzione di P&P;
- **trattamento degli ultracentenari:** analoga operazione di individuazione è stata fatta per gli ultracentenari;
- **prevalenza di una sola “identità” nei casi degli studenti lavoratori:** sono state scritte delle regole deterministiche di prevalenza, al fine di poter avere un solo comune di attrazione per la mobilità della persona;
- **lavoratori o studenti con residenza non verificata dal censimento:** le persone iscritte in anagrafe e non censite al momento del Censimento sono state circa 2,3 milioni. Il 75% di essi è presente in almeno uno degli archivi relativi alla Popolazione degli studenti e degli Universitari, o della Popolazione dei Lavoratori scelti per la costruzione del Sistema P&P.<sup>12</sup> Per questi microdati si è ritenuto opportuno verificare la residenza mediante informazioni amministrative ricavate dagli archivi di cui si è detto, e dalla base dati anagrafica del MEF.

<sup>12</sup> Questa evidenza consente di avvalorare l'ipotesi che il SIM (dove saranno armonizzati tutti gli archivi di P&P a regime), sia funzionale ad un uso di revisione anagrafica a cura degli enti titolari della tenuta dei registri di popolazione residente: segnalare persone che non presentano alcun legame con il territorio di competenza e quindi da candidare alla cancellazione dalla anagrafe; viceversa segnalare persone che, anche se non sono iscritte nell'anagrafe del Comune, presentano legami con il territorio di competenza e sono quindi da candidare all'iscrizione.



Le regole di editing deterministiche applicate sono:<sup>13</sup>

- se Domicilio Fiscale = Comune registrato in LAC;
- Comune di Residenza = Comune registrato in LAC;
- se Domicilio Fiscale  $\diamond$  Comune registrato in LAC , e Domicilio Fiscale = Comune di Lavoro;
- Comune di Residenza = Comune di Lavoro;
- se Domicilio Fiscale  $\diamond$  Comune registrato in LAC , e Domicilio Fiscale = Comune di Studio;
- Comune di Residenza = Comune di Studio;
- se Domicilio Fiscale  $\diamond$  Comune registrato in LAC , e Domicilio Fiscale = Comune Sede corso universitario.

Comune di Residenza = Comune sede del corso Universitario.

### 2.2.3 L'analisi della Integrabilità

#### a) Le criticità della Chiave di linkage

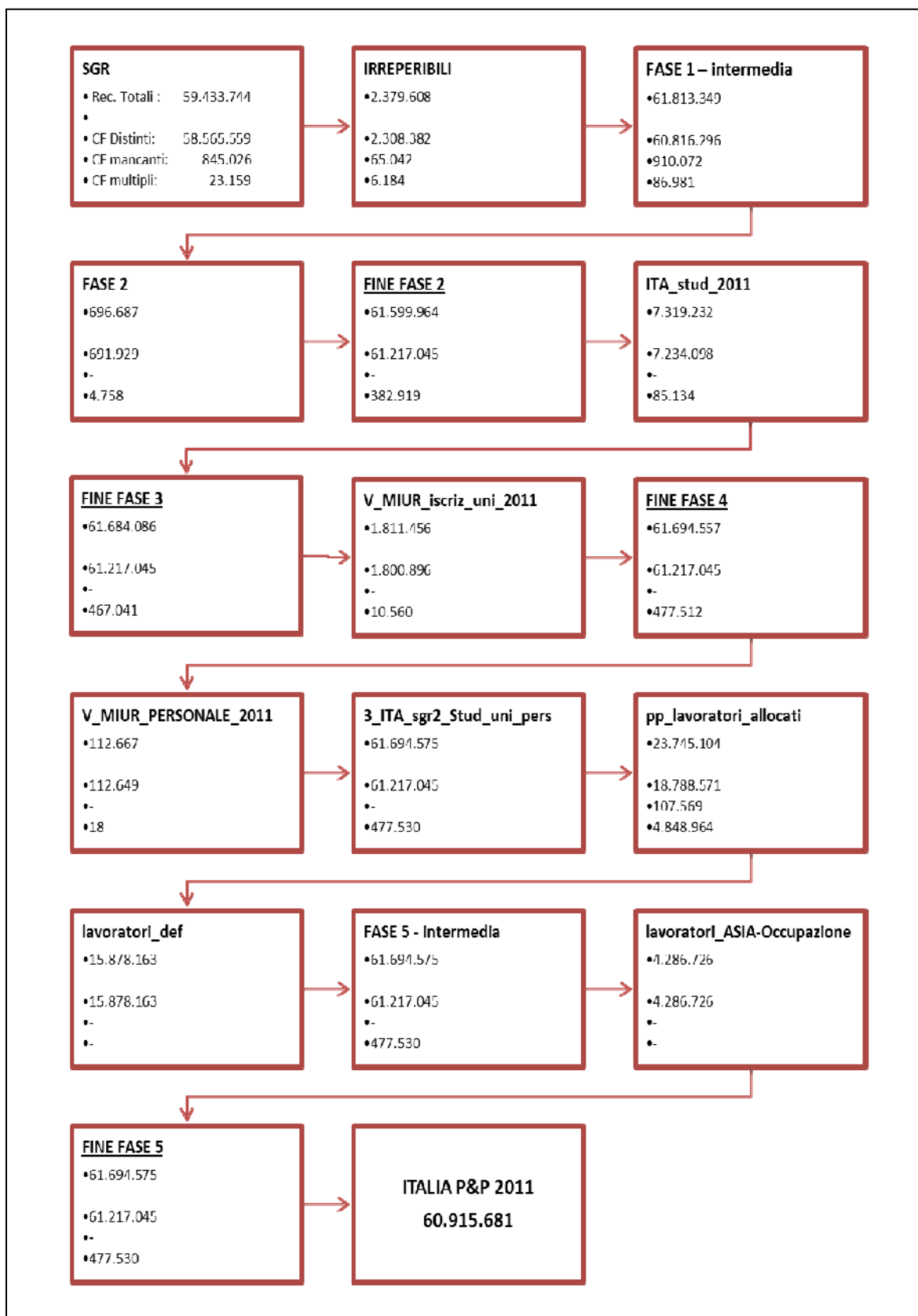
Il prototipo è stato costruito con un processo di *linkage* deterministico, seriale, senza recupero dei residui di ciascun passo, usando come chiave di linkage il Codice fiscale dell'individuo, perché omogenea e presente in tutte le fonti selezionate. Tale scelta – effettuata su microdati non integrati in SIM al momento della lavorazione, e dunque non “puliti” per ciò che concerne la loro identificazione – ha comportato un tasso di integrazione complessivo minore di quello potenziale a regime: nel prototipo, alla fine di ogni passo, sono stati accantonati in un serbatoio, come *casi non risolti*, tutti i microdati associati a codici fiscali nulli o multipli.

Nel prospetto seguente è riportato il dettaglio (i dati sono in valore assoluto, in milioni) per ogni passo, dei record totali, distinti, mancanti e multipli.

---

<sup>13</sup> La revisione delle persone non censite durante il censimento è iniziata a marzo 2012 ed è in corso di ultimazione. Il sistema P&P acquisirà i risultati definitivi appena disponibili.

**Prospetto 2.2 - Schema del processo, fasi e risultati**



**b) La qualità della concordanza della variabile Luogo di Studio (lavoro) tra fonti amministrative e fonte statistica (Censimento della popolazione)**

Per verificare la qualità della variabile amministrativa “Comune di lavoro/studio” del sistema P&P, si è scelta la strada della quantificazione della concordanza tra le modalità espresse nel sistema, e le modalità della omologa variabile censuaria.

Il confronto è stato effettuato con i dati ancora grezzi - per questa variabile - del Censimento della popolazione del 2011 (quesito 7.1<sup>14</sup> e 7.2<sup>15</sup>), riguardanti il motivo dello spostamento e il luogo di destinazione (Tavola 2.2).

**Tavola 2.2 - Unità nel sistema P&P e nel Censimento, e tasso di linkage tra le due fonti**

FONTE	Unità	
	Presenti nella fonte Valore assoluto	Tasso di linkage della fonte
Censimento	58.946.880	98,49
P&P	60.945.743	95,26

Confrontando la numerosità totale dei Codici fiscali distinti dalle due fonti si evince una differenza di circa 2 milioni di individui. Il risultato del linkage tra le due fonti risulta essere pari al 98,5% con riferimento al censimento e al 95,3% con riferimento al sistema P&P (Tavola 2.3): le unità in comune tra la base di dati P&P ed il Censimento sono quasi la totalità.

Tuttavia le numerosità delle diverse sottopopolazioni risultano distanti, come evidenziato nella tavola 2.3. Rispetto agli studenti i motivi sono da ricercare nella composizione degli archivi di partenza, da cui è stato ricavato il sistema P&P: infatti nelle fonti MIUR non sono presenti le scuole d’infanzia ed i corsi di formazione professionale; riguardo il lavoro, oltre alla considerazione che circa il 15% dei lavoratori non sono localizzati nel sistema P&P, si può supporre che parte delle differenze siano determinate dalla presenza, nel censimento, di lavoratori non regolari.

**Tavola 2.3 - Studenti e Lavoratori del censimento linkati nel sistema P&P**

TIPO DI FONTE E INDIVIDUO	Unità	
	Valore assoluto	Percentuale sul tipo individuo
Censimento		
- Studenti e Universitari	11.587.847	98,61
- Lavoratori	21.820.871	98,43
Totale	33.408.718	98,49
P&P		
- Studenti	6.649.637	97,53
- Universitari	1.719.389	97,68
- Lavoro	18.815.359	96,40
Totale	27.184.385	96,76

Infine, confrontando il luogo di lavoro/studio riportato nel censimento con quello nella base di P&P, risulta che la base di P&P è concorde per il 87% con il Censimento per lo studio, e per il 66% per il lavoro (Tavola 2.4): è quindi necessario affinare il metodo di localizzazione dei lavoratori utilizzato nella costruzione del prototipo di P&P per aumentare la quantità assoluta di lavoratori localizzati e la quantità relativa di casi concordi. Tuttavia ci si attende che la diversità delle fonti di origine dei dati non consentirà mai di azzerare le differenze.

<sup>14</sup> Quesito 7.1. Si reca giornalmente al luogo abituale di studio o di lavoro?: (1) Sì, mi reco al luogo di studio (compresi corsi di formazione professionale) [andare a dom. 7.2]; (2) Sì, mi reco al luogo di lavoro; (3) No, perché studio nel mio alloggio; (4) No, perché lavoro nel mio alloggio; (5) No, perché non ho una sede fissa di lavoro (piazziisti, rappresentanti, ecc.) [andare a dom. 8.1]; (6) No, perché non studio, non lavoro e non frequento corsi di formazione professionale.

<sup>15</sup> Quesito 7.2. Dove si trova il luogo abituale di studio o di lavoro?: (1) In questo comune; (2) In un altro comune italiano; (3) All’estero.

**Tavola 2.4 - Unità per le quali c'è concordanza sul comune di attività nel sistema P&P, e nel Censimento**

ATTIVITÀ SVOLTA NEL COMUNE	Valore assoluto	Percentuale su P&P	Percentuale su CENS
Studio	7.475.940	87,15	63,62
Lavoro	12.985.138	66,53	58,57

### 2.3 I risultati

Gli obiettivi previsti all'inizio delle attività (marzo 2013) erano i seguenti:

1. Costruire la matrice origini/destinazioni della mobilità quotidiana per Lavoro/Studio per comune utilizzando esclusivamente microdati amministrativi;
2. Quantificare all'interno della Popolazione Residente iscritta in Anagrafe le componenti (*city users*):
  - Resident city users;
  - Embedded city users;
  - Daily city users.
3. Quantificare la Popolazione Insistente sul territorio di ciascun comune.

Il risultato finale del sistema è stata la attribuzione del comune di lavoro/studio a tutti i lavoratori e gli studenti: anche di coloro che non erano stati reperiti dalla rilevazione censuaria o non erano iscritti in una anagrafe italiana.

I risultati della sperimentazione nel 2013 mostrano che utilizzando i dati amministrativi è possibile produrre la matrice OD<sup>16</sup> a livello comunale. Queste stesse fonti consentono anche di distinguere i cosiddetti<sup>17</sup> Residenti Stanziali (*Resident city users*) da quelli Non Stanziali, ma la sperimentazione ha anche dimostrato come le fonti amministrative non contengano sufficienti informazioni atte a distinguere tra i Residenti non stanziali i Pendolari giornalieri (*Daily city users*) dai Temporaneamente Dimoranti (*Embedded city users*).

Infatti le informazioni amministrative non permettono di determinare la “frequenza” con cui un individuo usa un determinato territorio e conseguentemente la “frequenza” della sua mobilità. Per raggiungere questo secondo obiettivo la sperimentazione proseguirà nel 2014 tentando di integrare i dati amministrativi disponibili con altre tipologie di informazioni (BIG DATA) e di sviluppare modelli di stima che utilizzano le distanze fra i luoghi.

In sintesi, gli obiettivi raggiunti sono stati i seguenti:

1. E' stata costruita la matrice origini/destinazioni per Lavoro/Studio per tutti i comuni italiani;
2. Sono state calcolate, all'interno della Popolazione Residente iscritta in anagrafe la popolazione dei *Resident city users* e quella che unisce le due tipologie di *Embedded city users* e *Daily city users*.

### 2.4 Prime ipotesi di prodotti

I prodotti sono stati elaborati al fine di rendere evidenti le potenzialità delle fonti integrate riguardo all'identificazione e analisi della popolazione insistente su un territorio.

#### **La matrice Origine Destinazione e le popolazioni derivabili**

Nella matrice sono contenute due tipologie di *city users*: i *Resident city users* e i (*Daily*

<sup>16</sup> Finora la matrice OD tra comuni della mobilità giornaliera per motivi di studio e lavoro è stata prodotta dall'Istat una volta ogni dieci anni usando le informazioni raccolte con il questionario di famiglia del censimento della popolazione e delle abitazioni.

<sup>17</sup> Cfr § 2.1.2.

+*Embedded*) city users, che aggregati opportunamente portano alla quantificazione della popolazione insistente sul territorio di un comune. I dati del sistema “Persons&Places” consentono anche di distinguere la mobilità per lavoro da quella per studio.

Nella tabella seguente è riportato un primo dato relativo alla sola componente stazionaria della popolazione insistente. La popolazione degli universitari si conferma essere quella più mobile sul territorio: 80 universitari su 100 sono “fuori sede”, ma anche 50 lavoratori su 100 escono dal comune di residenza per recarsi nella sede di lavoro.

**Tavola 2.5 - Popolazione con una attività di lavoro/studio in Italia, popolazione stazionaria, e tasso di stazionarietà. ITALIA**

ATTIVITÀ	Popolazione con una attività di lavoro/studio. ITALIA		
	Totale	Residenti statici	Tasso di stazionarietà
Lavoro	19.660.054	9.530.072	48,5
Università	1.500.464	308.441	20,6
Scuola	6.845.461	5.033.119	73,5
Totale	28.005.979	14.871.632	53,1

Di seguito (Tavola 2.6), per il Comune di Milano, viene calcolata la popolazione residente per destinazione di studio e lavoro e quella che insiste nel comune per origine. Dalla tavola si evidenzia come il sistema P&P, per il comune di Milano, considera 1.240 mila residenti, dei quali 145 mila lavorano o studiano al di fuori del territorio comunale (residenti in “uscita”). Gli individui che lavorano o studiano nel comune di Milano, pur essendo residenti in altri comuni, sono circa 640 mila (di cui l’83% sono lavoratori).

I dati mostrano come i “posti” di studio a lavoro *offerti* dal Comune siano più numerosi di quelli “occupati” dai residenti: 141 mila contro 114 mila per gli studenti; 141 mila contro 24 mila per gli universitari; 878 mila contro 348 mila per i lavoratori. In sintesi si può confermare che il Comune di Milano è una zona attrattiva dal punto di vista dei flussi di popolazione generati da attività legate al lavoro e allo studio (LS). Complessivamente la popolazione insistente per studio e lavoro è di oltre un milione (1.160 mila), mentre il numero degli occupati e degli studenti residenti a Milano (e indipendentemente dalla loro destinazione di studio a lavoro) è pari a di 632 mila.

**Tavola 2.6 - Popolazione che studia o lavora a Milano, per luogo di origine e destinazione**

Tipologia di residenti	Luogo di Residenza (origine)	Luogo di Studio/Lavoro (destinazione)				
		Comune	Entro la Provincia	Fuori Provincia	Fuori Regione	Totale studenti residenti a Milano
Studenti di Scuola Primaria e Secondaria	Comune	113.969	2.253	672	896	117.790
	Entro la Provincia	23.224	...	...	...	...
	Fuori Provincia	2.810	...	...	...	...
	Fuori Regione	999	...	...	...	...
	Totale "posti universitari" a Milano	141.002	...	...	...	...
						Totale universitari residenti a Milano
Universitari	Comune	24.362	257	972	1.262	26.853
	Entro la Provincia	27.918	...	...	...	...
	Fuori Provincia	53.661	...	...	...	...
	Fuori Regione	35.311	...	...	...	...
	Totale "posti studente" a Milano	141.252	...	...	...	...
						Totale lavoratori residenti a Milano
Lavoratori	Comune	347.954	67.846	27.959	43.543	487.302
	Entro la Provincia	202.643	...	...	...	...
	Fuori la Provincia	168.421	...	...	...	...
	Fuori Regione	159.572	...	...	...	...
	Totale "posti lavoro" occupati a Milano	878.590	...	...	...	...

**Indicatori**

La ricerca di opportune forme di diffusione dei dati ha consentito di individuare i primi possibili indicatori derivabili dal sistema “Persons&Places”. Tra di essi:

**Indice di autocontenimento:** rappresenta la capacità di contenere all’interno del proprio territorio la domanda di lavoro o studio dei residenti. Esso è il rapporto tra gli spostamenti interni a un’area e gli spostamenti generati nella stessa area (flussi che si esauriscono all'interno dell'area e flussi che escono dall'area stessa). Ad esempio nel caso degli studenti l’indice assume la seguente forma:

$$I_{aut}(STU) = (STU\_RES)/(STU\_OUT+STU\_RES)$$

Dove :

STU\_RES = studenti residenti nel comune i-esimo

STU\_OUT = flusso degli studenti in uscita dal comune i-esimo

STU\_IN = flusso degli studenti in entrata al comune i-esimo

**Indice di attrazione:** Indica la capacità di un'area di attrarre flussi dall'esterno al suo interno. La quantità dei flussi in ingresso è rapportata ai flussi generati dall'area (flussi che si esauriscono all'interno dell'area e flussi che escono dall'area stessa). Sempre nel caso degli studenti l'indice assume la seguente forma:

$$I_{\text{attr}}(\text{STU}) = (\text{STU\_IN})/(\text{STU\_IN}+\text{STU\_OUT}+\text{STU\_RES})$$

Nella tavola seguente sono riportati i gli indici di attrazione e autocontenimento calcolati per alcuni grandi comuni.

**Tavola 2.7 - Indici di attrazione e di autocontenimento per lavoratori per alcuni grandi comuni. Posti di lavoro localizzati all'interno del confine comunale, e Lavoratori residenti nel comune**

Comune	Indice di autocontenimento	Indice di attrazione	Posti di lavoro sul territorio comun.*	Lavoratori residenti nel Comune*
Roma	84,0	32,8	1.212	912
Palermo	81,3	32,4	207	161
Milano	71,4	60,4	879	487
Napoli	70,8	40,5	310	223
Firenze	68,9	43,1	198	136
Torino	67,1	36,5	388	311
Cagliari	67,0	53,0	84	46
Bologna	65,4	39,5	188	144

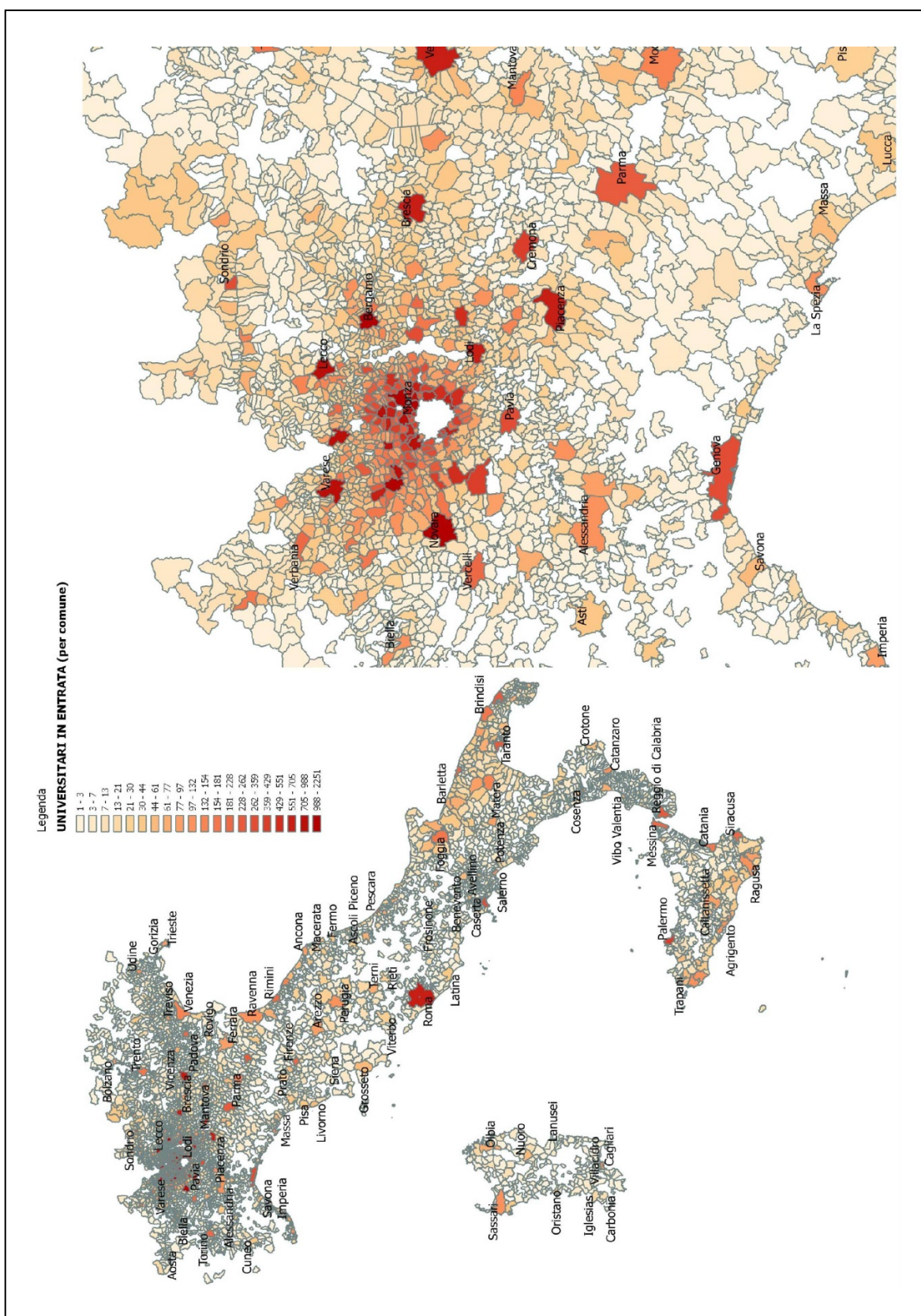
(\*) migliaia

### Rappresentazioni cartografiche

Di seguito vengono presentati alcuni cartogrammi<sup>18</sup> riguardanti i contingenti in “entrata” e in “uscita” per il Comune di Milano degli studenti universitari e dei lavoratori, con due dettagli differenti: uno per l'area di Milano e uno per l'Italia.

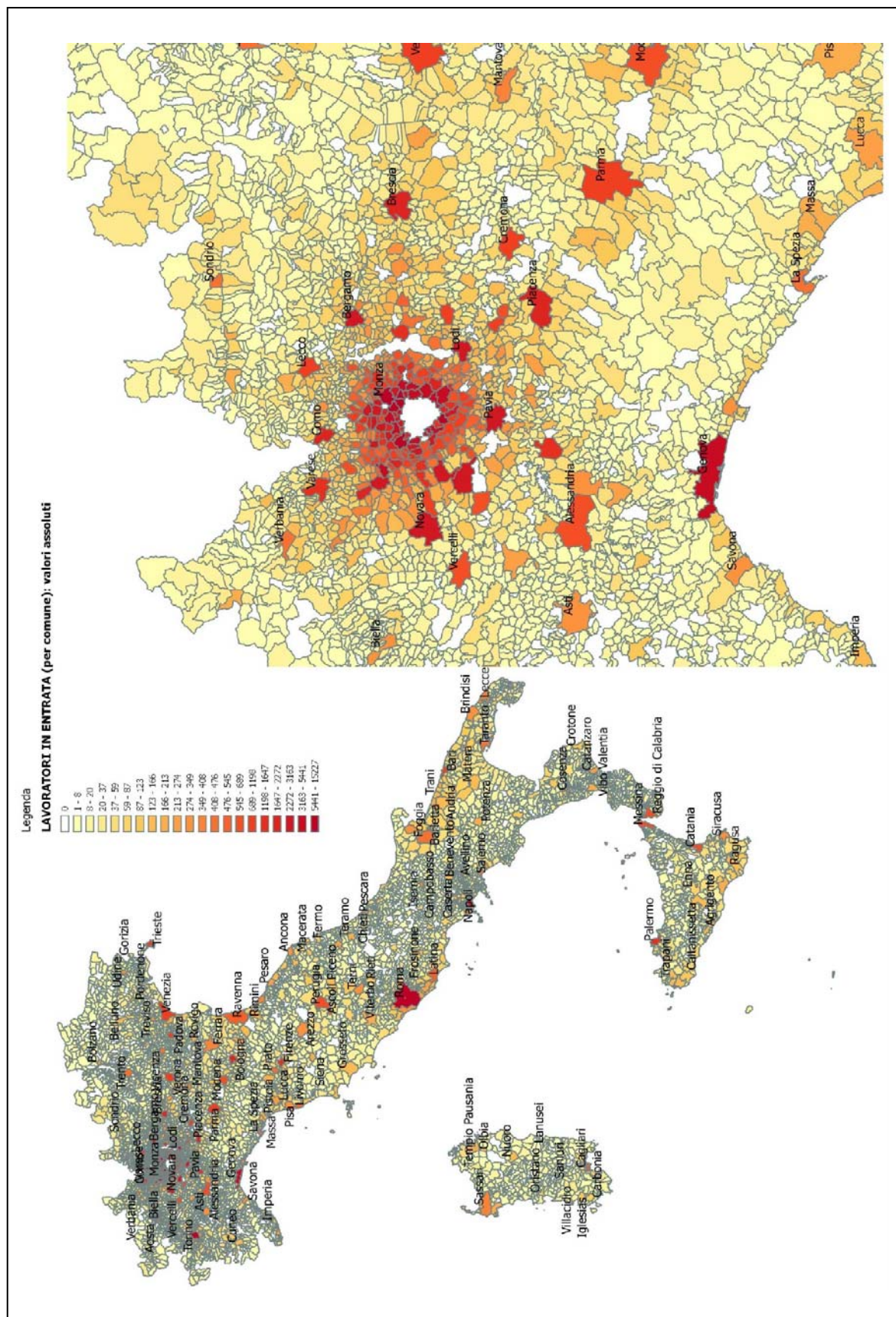
<sup>18</sup> Tutti i cartogrammi sono elaborati con QGIS 2.0.1; le classi di ampiezza sono state calcolate, in maniera automatica, utilizzando l'opzione Natural Breaks (Jenks) su 20 classi, basato su Jenks (1977).

Cartogramma 2.1 - Universitari in entrata per comune (Area Milano e Italia)





Cartogramma 2.2 – Lavoratori in entrata per comune: valori assoluti (Area Milano e Italia)



## 2.5 Sviluppi futuri

I risultati della sperimentazione nel 2013 mostrano che utilizzando i dati amministrativi è possibile produrre la matrice OD a livello comunale.

Queste stesse fonti consentono anche di distinguere i cosiddetti<sup>19</sup> Residenti Stanziali (*Resident city users*) da quelli Non Stanziali, ma la sperimentazione ha anche dimostrato come le fonti amministrative non contengano sufficienti informazioni atte a distinguere tra i Residenti non stanziali i Pendolari giornalieri (*Daily city users*) dai Temporaneamente Dimoranti (*Embedded city users*).

Un'ulteriore evidenza riscontrata dalla sperimentazione è la carenza di informazione sul luogo di lavoro per tutti quei lavoratori non rientranti nei settori coperti dal registro ASIA-Imprese e dal CIS 2011, (o rientranti nel campo di osservazione ma non allocati): per circa tre milioni di lavoratori la procedura sviluppata ad hoc dal progetto Archimede - proposta nel paragrafo 2.6 - Approfondimento - non è riuscita ad identificare il luogo di lavoro.

Pertanto, i principali sviluppi futuri riguarderanno:

1. Il miglioramento della qualità dell'informazione con particolare riguardo al luogo di lavoro degli individui. A questo fine da un lato sarà perfezionato l'uso delle informazioni desumibili da fonte INAIL, dall'altro lato si effettuerà un'analisi per l'individuazione di ulteriori fonti amministrative da integrare;
2. Lo sviluppo di metodologie per identificare le sottopopolazioni di individui che usano in maniera differente il territorio (distinzione tra *Embedded* e *Daily city users*). Per raggiungere questo secondo obiettivo la sperimentazione proseguirà nel 2014 tentando di integrare i dati amministrativi disponibili con altre tipologie di informazioni (BIG DATA) e di sviluppare modelli di stima che utilizzano le distanze fra i luoghi. La prima linea di ricerca, in collaborazione con il Dipartimento DIQR Istat, il Cnr e l'Università di Pisa, si fonderà sull'utilizzo di *Big data* per le analisi sulla mobilità, in particolare sull'uso della traccia dei dispositivi cellulari durante attività di chiamata (Call data record - Cdr). La seconda si fonderà sulla realizzazione di modelli probabilistici che utilizzano le "distanze" fra luogo di residenza e luogo di lavoro o studio, ed è condotta all'interno del Dipartimento;
3. L'ampliamento delle informazioni per due temi specifici: la mobilità familiare e lo studio dei bacini universitari.

---

<sup>19</sup> fr § 2.1.2 .

## Appendice. L'identificazione del luogo di lavoro per i lavoratori occupati che non rientrano nel campo di rilevazione del Censimento dell'Industria e Servizi 2011

Il sottoprogetto sperimentale “popolazione insistente” ha l'obiettivo di identificare e classificare i differenti user di un definito ambito territoriale. Il progetto prevede, per ciascun individuo, l'integrazione di informazioni sulla “residenza anagrafica” con quelle relative al luogo di studio (in scuole primarie, secondarie e università) e al luogo di lavoro.

L'input per il luogo di lavoro è dato dal Sistema ASIA-Occupazione, che però alla data attuale non integra alcune fonti e individua il luogo di lavoro (localizzazione dell'unità locale dell'impresa in cui presta la propria attività il lavoratore) solo per i “Dipendenti” e gli “indipendenti” che lavorano nel sottoinsieme delle imprese comprese nel campo di osservazione del CIS, nel mese di Dicembre 2011.

Ai fini della sperimentazione risulta quindi necessario individuare il luogo di lavoro per:

1. i lavoratori presenti in fonti non integrate nel DB-Occupazione: EX INPDAP, Autonomi dell'Agricoltura e Lavoratori Domestici;
2. i dipendenti, presenti nel DB-Occupazione ma occupati in Unità (Enti, Imprese, Associazioni, ...) esterne al campo di osservazione del CIS ovvero, tutti coloro che prestano attività lavorativa presso imprese non considerate in ambito CIS;
3. per tutti i parasubordinati e lavoratori interinali che svolgano la loro attività per Unità appartenenti o non al campo di osservazione del CIS.

A meno di casi particolari (ad esempio i Lavoratori Domestici la cui fonte nel mese di Dicembre presenta rapporti di lavoro mancanti a causa di un ritardo nella registrazione degli stessi) la localizzazione è stata identificata per il mese di Dicembre per garantire migliore allineamento con i dati derivati dal CIS.

Il presente documento descrive brevemente i metodi sviluppati e i risultati finali ottenuti.

### *Fonti non integrate nel Sistema ASIA-Occupazione*

#### *a) Autonomi dell'Agricoltura*

L'art. 2135 del codice civile (modificato con l'art. 1 del D.Lgs. 228/2001) precisa che: è imprenditore agricolo chi esercita un'attività diretta alla coltivazione del fondo, alla silvicoltura, all'allevamento di animali e attività connesse. La peculiarità dell'imprenditore agricolo, quindi, è data dal particolare contenuto del tipo di attività economica organizzata e mirata alla produzione di beni. Gli Autonomi dell'Agricoltura sono dunque: i coltivatori diretti e gli imprenditori agricoli professionali.

L'Archivio degli Autonomi dell'Agricoltura contiene i dati contributivi relativi ai soggetti sopra citati. A causa della specificità del settore per quel che riguarda il luogo di lavoro (che può essere caratterizzato da più terreni agricoli non contigui) e in assenza di un registro delle aziende agricole che permetta di impostare correttamente la problematica si è deciso di localizzare tutti gli Autonomi dell'Agricoltura nel luogo di residenza (472.580 lavoratori).

#### *b) Lavoratori Domestici*

L'Archivio dei Lavoratori Domestici contiene i dati contributivi relativi a tali soggetti.

Per la costruzione delle posizioni si considera il contratto di ogni lavoratore in essere nel mese di dicembre. A tale posizione si riporta il luogo di lavoro corrispondente presente nell'archivio.

Nella circostanza in cui il lavoratore domestico presenti rapporti di lavoro con più datori (24.838 casi) si è optato, per semplicità, per la selezione della prima posizione utile posseduta dal lavoratore. Considerando il mese di dicembre per i lavoratori domestici si ottiene che solo 471.529 codici fiscali distinti sono localizzati su un totale di 789.784. lavoratori. La causa è da attribuire alla completezza della fonte, ovvero nel mese di Dicembre la fonte presenta rapporti di lavoro mancanti a causa di un ritardo nella registrazione degli stessi. Per tale motivo si è optato

per la localizzazione dei lavoratori domestici nel mese di ottobre, per il quale risulta che 660.977 lavoratori sono localizzati su un totale di 700.919 codici fiscali distinti. Da questo punto in poi, nonostante i lavoratori domestici siano riferiti al mese di ottobre, si parlerà indifferentemente di lavoratori a dicembre.

Relativamente ai lavoratori domestici per i quali nei file INPS non si è potuto associare nessun comune di lavoro (informazioni *missing* sui datori di lavoro) ai fini di tale sperimentazione, non si è presa alcuna decisione in merito alla loro collocazione sul territorio, rimangono, quindi non localizzati 15.383 codici fiscali distinti.

### c) Posizioni assicurative EX INPDAP

Poiché per le posizioni assicurative ex-INPDAP non si dispone di unità locali, si è deciso di suddividere gli enti in “teoricamente” uni-localizzati quali ad esempio: comuni, amministrazioni provinciali e regionali, istituti scolastici etc.; e in enti “potenzialmente” pluri-localizzati.

Ogni lavoratore che afferisce al primo gruppo di enti è localizzato nel comune in cui ha sede l’ente di appartenenza. In questo caso i lavoratori localizzati sono 1.710.027 su un totale di 1.856.525, poiché per alcuni enti l’identificativo geografico è *missing*.

Per i lavoratori che afferiscono ad enti “potenzialmente” pluri-localizzati (1.295.238 individui) il procedimento di recupero dell’indirizzo, e quindi del comune di lavoro, passa per gli archivi: INAIL DNA<sup>20</sup> contenente i rapporti di lavoro e la Posizione Assicurativa Territoriale e INAIL PAT<sup>21</sup> contenente, invece, gli indirizzi corrispondenti ai codici PAT. Tale processo è descritto di seguito.

1. si effettua un primo aggancio con l’Archivio INAIL DNA, al fine di attribuire ad ogni lavoratore di un determinato ente un codice PAT;
2. successivamente si aggancia l’Archivio INAIL PAT tramite il codice PAT individuato per ogni lavoratore in modo da recuperare un indirizzo e quindi il relativo comune di lavoro.

Dopo aver agganciato il file Posizioni assicurative INPDAP con gli archivi INAIL DNA e INAIL PAT ne risulta che solo nel caso in cui il codice PAT è uguale al comune dell’ente il lavoratore risulta localizzato (367.148 individui), in tutti gli altri casi non è associata nessuna localizzazione al lavoratore. Tuttavia tale operazione può produrre una missclassificazione nei casi in cui ad un singolo ente con più unità locali venga assegnato un unico codice PAT (è questo, ad esempio, il caso dell’ISTAT).

Complessivamente il risultato finale è stato che i lavoratori appartenenti all’archivio Posizioni assicurative ex-INPDAP localizzati sono 2.065.920 su un totale di 3.153.323 codici fiscali distinti.

In relazione ai lavoratori ex-INPDAP per cui non si è potuto associare nessun comune di lavoro non si è presa alcuna decisione in merito rimangono quindi non localizzati 1.087.403 codici fiscali distinti.

<sup>20</sup> DNA (denuncia nominativa assicurati): Il decreto legislativo n. 38/2000 art. 14, comma 2 abolisce la DNA prevista per dipendenti e i parasubordinati. Dall’11 gennaio 2008 i datori di lavoro, in caso di assunzione, cessazione, trasformazione e proroga del rapporto di lavoro devono effettuare una sola comunicazione per via telematica ai Servizi per l’impiego. I datori di lavoro domestico (colf e badanti) dal 29 gennaio 2009 devono effettuare le comunicazioni obbligatorie per l’assunzione, la cessazione, la trasformazione e la proroga dei rapporti di lavoro all’INPS. Dal 18 agosto 2008 è obbligatoria la DNA all’INAIL per i collaboratori, i coadiuvanti e i soci sia artigiani che non artigiani, quando non sono oggetto della comunicazione preventiva di instaurazione del rapporto di lavoro ai sensi dell’art. 23, comma 8, DPR n. 1124/1965 (come modificato dall’art. 39, comma 8 del decreto legge n. 112/2008, convertito con modificazioni dalla legge n. 133/2008). Dall’11 gennaio 2008 i datori di lavoro pubblici e privati devono effettuare le comunicazioni di assunzione, cessazione, trasformazione e proroga dei rapporti di lavoro per via telematica ai Servizi per l’impiego (articolo 9 bis del decreto legge 1° ottobre 1996, n. 510, convertito, con modificazioni, dalla legge 28 novembre 1996, n. 608 e successive modificazioni). Per i datori di lavoro della Provincia autonoma di Bolzano l’obbligo decorre dal 1° dicembre 2008. (Fonte: INPS).

<sup>21</sup> PAT (posizione assicurativa territoriale): La posizione assicurativa territoriale (PAT) INAIL non corrisponde al numero delle unità locali in quanto, allo stato attuale, esistono situazioni in cui una stessa unità locale può avere diverse PAT che corrispondono ai diversi rischi assicurati (ad esempio, operai ed impiegati). Nell’archivio ditte ogni posizione assicurativa è ripetuta per ogni anno di competenza. In questo modo è possibile disporre delle informazioni sul numero di addetti per anno e correlare questa informazione con il numero di infortuni avvenuti. (Fonte: INPS).

**Tavola A.1 - Lavoratori della PA allocati, per tipologia di ente di appartenenza. Valori assoluti e proporzione sulla tipologia**

TIPOLOGIA DI ENTE	Lavoratori		
	Totale	Localizzati	Proporzione sulla tipologia
Agenzia delle entrate	29.057	5.533	19,04
Agenzia territoriale edilizia residenziale	6.578	4.666	70,93
Amministrazione provinciale	96.637	96.082	99,43
Amministrazione regionale	51.189	51.187	100
Azienda ospedaliera	237.850	122.769	51,62
Azienda sanitaria Locale	323.675	158.322	48,91
Azienda speciale	41.545	20.846	50,18
Camera di commercio	7.617	7.412	97,31
Comune	412.833	403.164	97,66
Conservatorio	6.896	6.821	98,91
Consorzio di enti locali	9.164	4.784	52,2
Ente regionale	20.846	8.727	41,86
Istituzione pubblica assistenza e beneficenza	39.071	36.954	94,58
Min. giustizia – procura	11.186	11.050	98,78
Min. giustizia – struttura penitenziaria	43.270	42.968	99,3
Min. giustizia – tribunale	21.983	21.647	98,47
MIUR – Scuola	1.024.343	1.001.914	97,81
MIUR – Ufficio scolastico regionale e provinciale	7.957	7.922	99,56
Struttura sanitaria territoriale	69.003	14.873	21,55
Unione di comuni	7.675	5.507	71,75
Altri enti	684.948	32.772	4,78
<b>TOTALE</b>	<b>3.153.323</b>	<b>2.065.920</b>	<b>65,52</b>

**d) Fonte: Asia-occupazione**

ASIA-Occupazione è una fonte prodotta internamente all'Istat, un registro costruito *ad hoc* dalla Direzione Centrale Archivi e Registri (Dcar) al fine di creare una struttura informativa per il Censimento virtuale CIS 2011. Esso è costituito da microdati integrati per unità economica (impresa) e individuo e contiene caratteri relativi all'unità (presenti in ASIA) e caratteri relativi all'individuo occupato.<sup>10</sup>

Per l'archivio Asia-Occupazione si selezionano i record presenti nell'archivio stesso che non sono stati allocati per il CIS 2011 considerando solo i contratti attivi nel mese di dicembre. Si tratta di 4.056.557 lavoratori da localizzare distinti in:

- lavoratori dipendenti non considerati in ambito CIS 2011;
- 1.037.014 parasubordinati;
- 487.094 interinali.

Si evidenzia che 695.212 individui lavorano presso enti che versano i contributi nelle casse dell'EX INPDAP e per i quali si è in possesso di una localizzazione. Nello specifico 549.680 dipendenti, 127.037 parasubordinati e 18.295 interinali. A tali individui è stato quindi associato l'identificativo geografico individuato per l'archivio Posizioni Assicurate EX INPDAP come descritto nel paragrafo precedente.

Per le imprese, inoltre si dispone delle unità locali in cui ogni impresa è dislocata. A tal punto, quindi, si procede ad agganciare i restanti 3.361.345 individui con l'archivio ASIA unità locali al fine di distinguere le imprese uni-localizzate da quelle pluri-localizzate. Da tale aggancio si ottengono i seguenti risultati.

- a) 1.539.047 lavoratori fanno riferimento a imprese uni-localizzate per cui risultano allocati nell'unità locale dell'impresa;
- b) I restanti 1.822.298 individui fanno riferimento a imprese con più unità locali per cui si procede con l'aggancio degli archivi INAIL DNA e INAIL PAT. Da tale aggancio ne risulta che nel ca-

so in cui il codice PAT è uguale al comune dell'unità locale, allora il lavoratore risulta localizzato (13.687 individui).

Al termine della suddetta procedura si nota che alcuni lavoratori possiedono più rapporti di lavoro anche con imprese diverse, a tal proposito, per poter assegnare ad ogni individuo un unico riferimento geografico si attribuisce ad ogni soggetto un unico rapporto di lavoro seguendo il criterio del contratto gerarchicamente superiore che segue, come deciso nell'ambito del sotto-progetto Precarietà lavorativa.

*Ranking* del contratto:

1. Dipendente a Tempo Indeterminato Full Time (TIF);
2. Dipendente a Tempo Indeterminato Part Time (TIP);
3. Lavoratori autonomi;
4. Altro dipendente;
5. Dipendente a Tempo Determinato (TD) e Lavoratori domestici;
6. Collaboratore e lavoratore interinale.

Eliminate le duplicazioni risulta che solo 1.856.853 codici fiscali distinti sono localizzati a dicembre 2011 su un totale di 4.056.557 codici fiscali distinti contenuti nel Asia-Occupazione per i quali non è stata attribuita una localizzazione nell'ambito del CIS 2011.

La tabella di frequenza dei lavoratori presenti nel Asia-Occupazione non allocati per il CIS 2001 risulta essere la seguente.

**Tavola A.2 - Lavoratori non localizzati del Sistema Informativo di ASIA-Occupazione, localizzati ex post, per tipologia lavorativa**

TIPOLOGIA	Lavoratori		
	con contratto attivo a dicembre e localizzati		
	Totale	Valore assoluto	Percentuale sul totale
Dipendenti	2.532.449	1.587.597	62,69
Interinali	487.094	23.465	4,82
Parasubordinati	1.037.014	245.260	23,65
<b>Totale</b>	<b>4.056.557</b>	<b>1.856.853</b>	<b>45,77</b>

### *Integrazione dei vari archivi, e risultati finali*

Riassumendo, si integrano insieme le seguenti fonti :

- la fonte Autonomi dell'Agricoltura con 472.580 "teste";
- l'archivio lavoratori domestici con 660.977 lavoratori localizzati;
- la fonte Posizioni Assicurative EX INPDAP con 2.065.920 dipendenti localizzati;
- la fonte ASIA-Occupazione parte non CIS 2011 con 1.856.853 lavoratori localizzati.

L'integrazione delle fonti in esame ha evidenziato l'esistenza di duplicazioni che sono state risolte come segue.

Primo criterio: contratto gerarchicamente superiore (come deciso nell'ambito del sotto-progetto Precarietà lavorativa).

*Ranking* del contratto:

1. Dipendente a Tempo Indeterminato Full Time (TIF);
2. Dipendente a Tempo Indeterminato Part Time (TIP);
3. Lavoratori autonomi;
4. Altro dipendente;
5. Dipendente a Tempo Determinato (TD) e Lavoratori domestici;
6. Collaboratore e lavoratore interinale.

Secondo criterio: fonte di provenienza gerarchicamente superiore (come deciso nell'ambito del sotto-progetto Precarietà lavorativa).

*Ranking* della fonte:

1. ASIA\_Occupazione;
2. INPDAP;
3. Autonomi in Agricoltura;
4. Lavoratori Domestici.

Nel caso di posizioni duplicate prima di ridurle a chiave a uno (codice fiscale lavoratore) si costruisce un *flag* che indica se il lavoratore era presente in più fonti e un flag che indica se il lavoratore era pluri-localizzato fra le fonti al fine di poter condurre ulteriori approfondimenti in seguito.

In conclusione, nella tavola seguente sono riportati i risultati del processo, al netto delle duplicazioni.

**Tavola A.3 - Lavoratori non compresi nel Sistema Informativo di ASIA-Occupazione (\*), per tipologia lavorativa**

Tipologia	Lavoratori		
	<i>con contratto attivo a dicembre e localizzati</i>		
	Totale	Valore assoluto	Percentuale sul totale
Dipendenti	5.730.104	4.033.751	70,40
Interinali	479.938	20.147	4,20
Parasubordinati	1.024.297	232.828	22,73
Totale	7.234.339	4.286.726	59,26

(\*) Persone occupate nella Pubblica amministrazione, occupate in settori non censiti dal Censimento dell'Industria e servizi, occupate come lavoratori domestici, occupate come autonomi in agricoltura.

### 3. Il sotto-progetto sperimentale: precarietà lavorativa<sup>22</sup>

#### 3.1 Le caratteristiche della sperimentazione

Il fenomeno del precariato lavorativo si presenta complesso e articolato in termini sia normativi, sia definitivi. Sebbene il principale problema del mercato del lavoro sia rappresentato oggi in Italia dalla disoccupazione, la presenza di posti di lavoro non stabili resta comunque un tema al centro di numerose analisi, sia economiche che sociali.

Il lavoro precario è un concetto che riguarda sia aspetti oggettivi della condizione lavorativa di un individuo, sia la percezione dei riflessi che la propria situazione occupazionale ha sulle scelte di vita. Visti gli impatti sociali soggettivi e dato che la normativa sul lavoro in Europa è ben differenziata, i confini di ciò che è o non è lavoro precario sono piuttosto sfocati. A livello internazionale non esiste una definizione formalizzata di "lavoro precario" e il dibattito è apertissimo.<sup>23</sup> Alcune raccomandazioni si evincono dai testi dell'ILO,<sup>24</sup> ma non si tratta di definizioni dai connotati statisticamente circoscritti; infatti, viene genericamente definito precario quel rapporto di lavoro in cui la sicurezza del lavoro, principale elemento del contratto di lavoro, è limitata.

##### 3.1.1 Obiettivi

Tenendo conto delle forme di lavoro riconosciute dall'ordinamento, e sulla base di fonti prevalentemente previdenziali e fiscali, è stato adottato un approccio al fenomeno della "precarietà lavorativa" incentrato sui concetti di flessibilità e atipicità (o lavoro non-standard). Inoltre,<sup>25</sup> è stato utilizzato il concetto di professionalità che ha indotto a circoscrivere ulteriormente l'individuazione dei contratti di lavoro precari tra quelle categorie contrattuali poste a cavallo tra il lavoro dipendente e quello autonomo. Risulta evidente come i criteri individuati siano limitativi poiché non integrano concetti quali il reddito, la formazione o informazioni quali la presenza di cassa integrazione di lungo periodo e le pluricommitenze. L'integrazione di ulteriori elementi nell'identificazione e classificazione dei lavoratori precari, solo parzialmente avviata nella fase sperimentale del progetto, potrà essere completata in una fase successiva con il fine sia di misurare il "grado" di precarietà (ad es. durate contrattuali o aspetti reddituali), sia nell'ottica di escludere parte degli individui dalla condizione di precarietà, o includerne altri normativamente inseriti in categorie contrattualmente stabili.

L'obiettivo principale del progetto è di effettuare **l'analisi di un universo di individui definiti "lavoratori precari", fotografandone aspetti socio-demografici e caratteri economico-lavorativi in un arco temporale, e seguirne le trasformazioni nel tempo.**

L'idea quindi che sta alla base del sotto-progetto non è quella di voler spiegare il fenomeno della precarietà lavorativa, né tantomeno capirlo e prevederne le variazioni temporali, bensì è quella di descriverlo ed osservarne i mutamenti nel tempo.

##### 3.1.2 Unità statistiche, popolazione di riferimento e principali definizioni adottate

I caratteri definitivi del fenomeno sono stati studiati a partire dall'analisi delle norme,<sup>26</sup> nonché dall'esame di indagini e sistemi informativi implementati sia all'interno che all'esterno dell'Istat. In sintesi: (a) il quadro normativo esaminato ha permesso di identificare le forme contrattuali di lavoro previste dall'ordinamento; (b) le indagini statistiche analizzate hanno fornito conoscenze sulle informazioni raccolte riguardo le attività lavorative degli individui; (c) i sistemi informativi esplo-

<sup>22</sup> I §§ 3.1, 3.2.1, 3.2.2, 3.3 e 3.4 sono a cura di D. Ercolani, il § 3.2.3 è a cura di M.D. De Gabrieli.

<sup>23</sup> 8° European Workshop on Labour Force Survey LFS Methodology. GUS Poland, Gdańsk. 23-24 May 2013: [http://www.stat.gov.pl/lfs2013/papers\\_and\\_presentations.htm](http://www.stat.gov.pl/lfs2013/papers_and_presentations.htm).

<sup>24</sup> ILO Thesaurus, 2011: "Precarious employment is a work relation where employment security, which is considered one of the principal elements of the labour contract, is lacking. This term encompasses temporary and fixed term labour contracts, work at home and subcontracting".

<sup>25</sup> Rapporto ISTAT 2013. Capitolo 3 - Il mercato del lavoro tra minori opportunità e maggiore partecipazione.

<sup>26</sup> Al fine di individuare il campo di osservazione della tematica in analisi è stata necessaria una ricognizione sulle recenti normative del mercato del lavoro per individuare le forme contrattuali presenti nel nostro ordinamento a partire dall'anno 2011.



rati hanno agevolato la ricognizione delle fonti informative disponibili e utilizzabili nel progetto.

Nello specifico sono stati analizzati i seguenti ambiti di riferimento:

- Legge Biagi sul mercato del lavoro (D.Lgs. 276/2003) e successive riforme o Leggi Finanziarie in tema di lavoro;
- Indagine Istat “EU-SILC” (Regolamento UE n. 1177/2003);
- Indagine Istat “Rilevazione Continua sulle Forze di Lavoro” (Regolamento UE n. 577/98);
- Indagine Isfol PLUS;<sup>27</sup>
- Sistema Integrato di Microdati (SIM) Istat.<sup>28</sup>

La letteratura consultata conferma che la precarietà lavorativa è considerata un concetto multidimensionale<sup>29</sup> in sintesi circoscrivibile alle seguenti dimensioni: sicurezza (durate contrattuali, diritti sindacali); *welfare* (sistema previdenziale); formazione; guadagno. Tuttavia, poiché solo alcune caratteristiche sono misurabili attraverso le fonti amministrative disponibili, si è posta l’attenzione sugli aspetti lavorativi osservabili nel nostro mercato del lavoro.

Alla luce della documentazione relativa a studi su aspetti terminologici legati al mercato del lavoro,<sup>30</sup> con il termine di *flessibilità* si intende il processo di progressivo e reciproco adattamento tra le esigenze individuali dei lavoratori e le istanze datoriali: da un lato sono flessibili i comportamenti che modulano il tempo di lavoro ceduto all’impresa attraverso istituti come il part-time (per il versante lavoratori) o il lavoro straordinario (sul versante imprese), dall’altro si fa riferimento al complesso dei contratti di lavoro atipici, ovvero a ciò che differisce dall’occupazione standard (lavoro dipendente a tempo indeterminato a tempo pieno e lavoro autonomo).<sup>31</sup> D’altro canto, dal punto di vista soggettivo, si può affermare che *la precarietà lavorativa rappresenta uno stato di incertezza rispetto al futuro occupazionale di un individuo data la presenza di caratteristiche di flessibilità che riducono l’orizzonte di pianificazione delle scelte abitative, riproduttive, di consumo o risparmio dell’individuo*. Di conseguenza, non potendo misurare la volontarietà della scelta di un lavoratore di modulare il proprio tempo di lavoro nell’ambito dell’attività lavorativa svolta, il lavoro precario è stato qui individuato tra quelli appartenenti all’insieme dei contratti di lavoro cosiddetti atipici, senza considerare aspetti della precarietà legati alla percezione soggettiva del disagio; ci si è posti bensì nell’ottica di osservare quelle caratteristiche del lavoro legate ad aspetti normativi ed operativi delle forme contrattuali che si traducono inevitabilmente in aumento della sensazione di incertezza.

Pertanto, la popolazione di riferimento è costituita dagli individui che al periodo di riferimento in analisi sono titolari di uno o più contratti di lavoro, anche svolti contemporaneamente.

E’ stato considerato lavoro precario: (a) il lavoro dipendente a termine; (b) il lavoro prestato in base a rapporti di collaborazione per svolgere attività aventi caratteristiche di subordinazione; (c) il lavoro autonomo consistente in attività aventi caratteristiche di subordinazione.

Per completezza, vale la pena focalizzare l’attenzione su alcuni ulteriori punti:

1. seppur classificati come lavoro flessibile, non tutti i contratti di lavoro atipico si configurano come lavoro precario (ad es. i contratti di lavoro a tempo indeterminato a tempo parziale

<sup>27</sup> PLUS - Participation Labour Unemployment Survey - è una rilevazione campionaria nazionale condotta dall’Isfol dal 2005 che produce indicatori relativi a redditi (da lavoro e familiari), istruzione e background familiare degli individui, servizi presenti sul territorio, salute. La completa copertura della popolazione e in particolare di tutti gli occupati e la struttura longitudinale dell’indagine permette di compiere analisi di flusso tra le diverse condizioni (non solo occupazionali) consentendo così l’analisi delle “storie lavorative” individuali. Vedi anche Mandrone, Radicchia (2012).

<sup>28</sup> E’ una infrastruttura statistica di base nella quale vengono integrati dati riferiti a individui, famiglie e unità economiche provenienti da una pluralità di basi dati di fonte amministrativa. Include dati relativi a: (1) caratteristiche di individui e famiglie; (2) caratteristiche delle unità in cui gli individui realizzano le loro attività (o la loro vita); (3) caratteristiche dei luoghi di residenza, lavoro e studio; (4) tipologie di relazioni tra individui, unità e luoghi.

<sup>29</sup> Working Lives Research Institute. Faculty of Social Sciences and Humanities - London Metropolitan University, “Study on Precarious work and social rights”, carried out for the European Commission. FINAL REPORT, April 2012.

<sup>30</sup> Mandrone (2008).

<sup>31</sup> È lavoratore autonomo - art. 2222 del Codice Civile - colui che si obbliga a compiere verso un corrispettivo un’opera o un servizio, con lavoro prevalentemente proprio e senza vincolo di subordinazione nei confronti del committente.

- lavoro parzialmente standard - volontariamente richiesti, oppure contratti a termine applicati ad amministratori, revisori o liquidatori di società);
- 2. tra i contratti di lavoro autonomo, è possibile osservare situazioni anomale di subordinazione (ad es. le P.IVA con mono committenza);
- 3. le definizioni sul precariato lavorativo adottate tendono ad escludere dal campo di osservazione quegli individui occupati stabilmente in imprese in crisi o i professionisti esposti ad una domanda di mercato il cui futuro è incerto indipendentemente dalla forma contrattuale: a tal proposito, non sono stati considerati lavoratori precari i cassa-integrati, i quali vengono configurati come occupati stabili con un grado di sicurezza non elevato in quanto esposti al rischio di espulsione dal mercato del lavoro;
- 4. nonostante alcune esperienze lavorative si configurino come periodo di formazione (ad es. dottorati o assegni di ricerca post-lauream), esse verranno valutate in fase sperimentale come attività lavorativa precaria in quanto il loro utilizzo ha di consuetudine finalità lavorative.<sup>32</sup>

In conclusione, **viene qui considerato lavoratore precario l'individuo che, in un determinato periodo di tempo, svolge esclusivamente lavori definiti "precari".**

### 3.1.3 Le fonti dei dati

Come detto, l'unità statistica del sotto-progetto è il lavoratore, il quale viene classificato in base alla sua condizione rispetto al fenomeno della precarietà lavorativa. L'obiettivo di osservare le transizioni da una condizione all'altra fa emergere l'esigenza di descrivere la condizione di uscita dall'occupazione (o condizione precedente a quella di occupato). Deve quindi essere possibile, relativamente a ciascun periodo di tempo considerato, classificare l'individuo come appartenente alle seguenti macro categorie lavorative:

1. lavoratore, precario (oggetto principale dell'analisi) o stabile (dipendente, autonomo);
2. altra condizione (ad esempio perdita di lavoro, pensionamento, decesso, emigrazione).

La prima categoria afferisce all'occupato sul quale vengono osservate caratteristiche lavorative, incluse quelle relative al suo eventuale datore di lavoro. La seconda categoria – contraddistinta dall'assenza di una attività lavorativa - è individuata a partire dall'analisi di fonti dai contenuti essenzialmente demo-sociali. Le fonti individuate (di natura amministrativa, censuaria, registri statistici e sistemi informativi), classificate per contenuti e condizioni osservabili, sono esposte nel seguente schema.

32 Altre categorie non sono state considerate nella sfera del precariato: i lavoratori collocati (iscritti alle liste) in mobilità e i Lavoratori Socialmente Utili, in quanto classificati come soggetti sottoposti a trattamenti previdenziali o assistenziali; i tirocini formativi e di inserimento lavorativo, in quanto sottoposti a formazione non soggetta a versamento di contributi previdenziali (normativamente non occupati).

**Prospetto 3.1 - Le fonti amministrative per il progetto precarietà lavorativa**

Fonte dei dati	Contenuti informativi	Condizioni osservabili
ASIA – Occupazione	Tipologia occupati / qualifica / durata contratto; luogo di lavoro	Relative al lavoratore
Ex INPDAP – Posizioni Assicurative	Tipologia occupati / qualifica / durata contratto; luogo di lavoro	Relative al lavoratore
Archivio Lavoratori Domestici	Tipologia / durata contratto; luogo di lavoro	Relative al lavoratore
Autonomi Agricoltura	Tipologia di occupati	Relative al lavoratore
MEF – Banca Dati Redditi: Modelli Unico/730/770 / Dichiarazioni IVA	Redditi; individuazione P.IVA monocommittenti	Relative al lavoratore
INAIL – DNA / PAT	Luogo di lavoro	Relative al lavoratore
Censimento Popolazione	Titolo di studio	Altra condizione
Liste Anagrafiche Comunali (LAC)	Caratteristiche anagrafiche dei componenti; dimensione familiare e tipo di famiglia	Altra condizione
Casellario dei pensionati	Individuazione del pensionato; percettori di pensioni assistenziali	Altra condizione
Gestione fiscale per prestazioni non pensionistiche	Indennità di disoccupazione / Mobilità / LSU-LPU	Altra condizione
MIUR - Anagrafe degli studenti delle scuole / Anagrafe studenti universitari / AFAM	Iscrizione	Altra condizione
Cause di morte	Individuazione persone decedute	Altra condizione
AIRE - Anagrafe della popolazione Italiana Residente all'Estero	Individuazione persone emigrate	Altra condizione
ASIA – Imprese	Caratteristiche impresa	Relative al datore di lavoro
Lista S13 – PA	Caratteristiche ente	Relative al datore di lavoro
Conto Annuale IGOP-RGS	Caratteristiche ente	Relative al datore di lavoro
Censimento PA / Non Profit	Caratteristiche ente	Relative al datore di lavoro

Ai fini del raggiungimento degli obiettivi previsti, il contenuto informativo più rilevante è rappresentato dalla fonte **ASIA-Occupazione**.<sup>33</sup> si tratta di una base di microdati integrati per impresa e occupato, costruita ai fini dell'implementazione del CIS (Censimento Industria e Servizi 2011), nonché dell'aggiornamento del Registro Istat ASIA - Imprese attive,<sup>34</sup> la creazione di questa fonte è stata basata su un set sperimentale di dati relativi all'annualità 2010. Il risultato è un prodotto complesso, dagli obiettivi versatili e di elevata qualità statistica, e per tale motivo rappresenta la fonte *core* del lavoro "Precarietà lavorativa" del progetto ARCHIMEDE. Nello specifico, ciascun individuo è stato classificato in base alla tipologia occupazionale all'interno dell'impresa con la quale ha avuto un rapporto di lavoro. La struttura integrata<sup>35</sup> contiene i principali aggregati occupazionali (posizioni lavorative e individui), le caratteristiche del rapporto di lavoro, inclusa l'osservazione di ammortizzatori sociali quali CIG (Cassa Integrazione Guadagni) e contratti di solidarietà, nonché alcune variabili demografiche del lavoratore.

<sup>33</sup> È costituito dall'integrazione complessa di un insieme di fonti amministrative, dove l'unità di analisi (unità economica impresa) è identificata sulla base di definizioni condivise anche a livello europeo, e l'occupazione è considerata un attributo di tale unità. La struttura informativa si compone di tre livelli informativi (impresa, lavoratori e relazioni tra lavoratori e imprese) l'attività lavorativa viene svolta. Vedi anche: Istat (2013).

<sup>34</sup> Archivio statistico delle imprese attive: rappresenta le imprese che hanno svolto un'attività produttiva nell'anno ed è il risultato dell'integrazione di numerose fonti amministrative. La produzione dell'archivio riguarda le unità economiche attive (imprese, UL) e i relativi caratteri anagrafici (denominazione) e di struttura (attività economica, dimensione, forma giuridica), nonché di localizzazione e demografici.

<sup>35</sup> Fonti integrate in ASIA Occupazione: (1. Lavoratori dipendenti) INPS – Emens/DMAG/CIG pagamento diretto/Ex ENPALS Sport e Spettacolo; (2. Lavoratori indipendenti) CCIAA – Persone con cariche d'impresa/ Soci di impresa; INPS – Imprenditori/Artigiani/Commercianti/Gestione Separata Parasubordinati; Agenzia delle Entrate – Modello unico; (3. Occupati esterni) INPS – Gestione Separata Parasubordinati; INAIL – Somministrazione di lavoro.

### 3.1.4 Le attività sperimentali

In relazione ai vincoli oggettivi che riguardano (1) la disponibilità delle fonti, (2) il grado di sviluppo dei sistemi informativi presenti in Istituto e (3) le attività di processo che le risorse umane messe in campo nel progetto sono state in grado di eseguire, le attività svolte durante il primo periodo di sperimentazione possono essere così sintetizzate:

- **fonti di dati selezionate:** sono state esplorate le fonti presenti in SIM relative all'occupazione;
- **periodo di riferimento:** sono stati estratti dati relativi alle annualità 2010 e 2011 e osservate le informazioni relative agli occupati nei due mesi di ottobre;<sup>36</sup> la riproduzione delle identiche informazioni per i due periodi successivi ha permesso di analizzare le trasformazioni del fenomeno per confronto tra stock alle due date;
- **popolazione di riferimento:** è stato analizzato l'universo degli occupati indipendenti e dipendenti di datori di lavoro / imprese residenti su tutto il territorio nazionale;
- **output realizzato:** set di microdati di lavoratori;
- **contenuti analizzati:** caratteristiche anagrafiche, lavorative e aspetti reddituali;
- **analisi effettuate:** analisi trasversali, analisi longitudinali, analisi dei flussi di entrata/uscita dalla condizione di occupato e analisi dei redditi.

## 3.2 I risultati

L'esito della sperimentazione è stato la predisposizione di insiemi di informazioni sui lavoratori opportunamente identificati nel periodo riferimento prescelto, la cui analisi ha permesso di classificare il lavoratore precario, nonché costruire tavole sulla base di variabili di natura economico-lavorativa e socio-demografica.

### 3.2.1 I dati elementari: struttura e contenuti informativi

Le fonti di microdati sul lavoratore alla base della sperimentazione hanno natura fiscale e previdenziale, e sono tutte presenti nel SIM:

- ASIA-Occupazione;
- posizioni Assicurative Ex INPDAP;
- lavoratori Autonomi in Agricoltura;
- lavoratori Domestici.

Queste fonti sono state studiate con lo scopo di estrarre da ognuna informazioni standardizzate coerenti con il processo di analisi integrata previsto.

Il principale contenuto informativo dell'integrazione è rappresentato dalla fonte denominata **ASIA-Occupazione**, già descritta in precedenza.

La seconda fonte di dati integrata, per numerosità di individui, è rappresentata dalla banca dati implementata dall'INPS - Gestione Ex INPDAP, denominata **Posizioni Assicurative Ex INPDAP**. Essa copre la quasi totalità del settore occupazionale della Pubblica Amministrazione.<sup>37</sup> La Posizione Assicurativa dell'iscritto contiene le seguenti informazioni: dati anagrafici e di residenza dell'iscritto; dati identificativi dei datori di lavoro; lo stato dei servizi prestati e le relative retribuzioni percepite; i periodi riconosciuti con l'emissione di un provvedimento (riscatti, ricingiunzioni, computi); le maggiorazioni di servizio. Concretamente, la banca dati delle Posizioni Assicurative è popolata dai seguenti soggetti: gli enti e le amministrazioni pubbliche datori di lavoro, gli iscritti, i patronati, gli operatori dell'Inps - Gestione Dipendenti Pubblici.

<sup>36</sup> Il mercato del lavoro è soggetto al fenomeno della c.d. "stagionalità", in particolar modo per alcuni settori di attività e con riguardo agli eventi di assunzione. Secondo la letteratura sul mercato del lavoro il mese di ottobre è tra quelli meno soggetti alla stagionalità delle assunzioni.

<sup>37</sup> Affrontare lo studio delle classificazioni dei settori di attività risulta al momento prematuro. Difatti, il settore ATECO è ben definito per le imprese incluse nel Registro ISTAT ASIA. Per altre fonti, invece, come ad esempio le Posizioni Assicurative Ex INPDAP, è tuttora in corso l'analisi che permetterà di classificare gli enti iscritti, e le loro unità organizzative/locali, in base alle tipologie istituzionali contemplate nella Lista S13 delle Pubbliche Amministrazioni.

La fonte sui **Lavoratori Autonomi in Agricoltura** si basa sull'art. 2135 del codice civile (modificato con l'art. 1 del D.Lgs. 228/2001) secondo il quale è imprenditore agricolo chi esercita un'attività diretta alla coltivazione del fondo, alla silvicoltura, all'allevamento di animali e attività connesse. La peculiarità dell'imprenditore agricolo, quindi, è data dal particolare contenuto del tipo di attività economica organizzata e mirata alla produzione di beni. Gli Autonomi dell'Agricoltura sono dunque: i coltivatori diretti e gli imprenditori agricoli professionali. L'archivio dei Lavoratori Autonomi in Agricoltura contiene i dati contributivi relativi ai soggetti sopra citati che soddisfano il combinarsi dei seguenti requisiti oggettivi: il fabbisogno lavorativo necessario per la gestione dell'azienda non deve essere inferiore a 104 giornate annue ed il nucleo coltivatore diretto deve far fronte autonomamente ad almeno un terzo del fabbisogno lavorativo annuo occorrente per la gestione dell'azienda.

La fonte sui **Lavoratori domestici** contiene dati e informazioni su coloro che prestano un'attività lavorativa continuativa per le necessità della vita familiare del datore di lavoro come ad esempio colf, assistenti familiari o *baby sitter*, governanti, camerieri, cuochi ed altri operatori connessi; rientrano in queste categorie anche i lavoratori che prestano tali attività presso comunità religiose (conventi, seminari), presso caserme e comandi militari, nonché presso le comunità senza fini di lucro, come orfanotrofi e ricoveri per anziani, il cui fine è prevalentemente assistenziale. L'archivio INPS dei Lavoratori Domestici contiene i dati contributivi relativi a tali soggetti.

Come meglio descritto in seguito, le fonti sopra descritte sono state analizzate, al fine di estrarre, elaborare e ottenere i seguenti output:

### **Output di tipo 1**

Corrisponde al livello di singola attività lavorativa svolta dal lavoratore presso un datore di lavoro nel periodo di riferimento preso in esame: mese di ottobre, annualità 2010 e 2011. L'output è rappresentato da un set di microdati costituito da record di variabili chiave (Anno - Lavoratore – Datore – Tipologia lavorativa), corredato da informazioni specifiche sulla tipologia lavorativa svolta, dove:

- Anno: annualità di riferimento delle fonti integrate;
- Lavoratore: codice fiscale lavoratore;
- Datore: codice fiscale datore di lavoro;
- Tipologia lavorativa: tipologia contrattuale secondo la quale il lavoratore svolge l'attività lavorativa presso il datore. Convenzionalmente, anche in relazione alle tipologie contrattuali osservabili, le tipologie di lavoro individuate sono costituite da:
  - Dipendente a tempo indeterminato;
  - Autonomo;
  - Altro dipendente;<sup>38</sup>
  - Dipendente a tempo determinato;
  - Collaboratore;
  - Lavoratore domestico.

Il precedente elenco rispetta un ordine ipotetico che va dalla condizione più stabile alla condizione più precaria.

### **Output di tipo 2**

Corrisponde al livello di singolo lavoratore, ovvero rappresenta la tipologia lavorativa principale svolta dal lavoratore durante il periodo di riferimento allo studio. L'output è un set di microdati costituito da record di variabili chiave (Anno - Lavoratore – Tipologia lavorativa principale), dove, nel caso di attività di lavoro multiple nel periodo di riferimento, la tipologia lavorativa principale è

<sup>38</sup> Questioni legate alla completezza e alla qualità dei dati amministrativi non hanno permesso di distinguere la determinatezza della scadenza contrattuale per numerosità scarse di lavoratori dipendenti. Di conseguenza è stato scelto di mantenere la modalità di tipologia di lavoro "altro dipendente" in quanto la presenza della scadenza contrattuale discrimina l'appartenenza dell'individuo alla classe dei lavoratori precari.

individuata in base al *ranking* sopra esposto. Il set di dati contiene informazioni anagrafiche connesse al lavoratore e l'individuazione della condizione di precariato costruita nel rispetto della regola secondo cui è lavoratore precario l'individuo che, nel periodo di tempo allo studio, svolge esclusivamente lavori definiti "precari". Il seguente prospetto sintetizza le condizioni di precarietà lavorativa ricostruite con l'implementazione dell'output di tipo 2 (cfr. § 3.2.2).

**Prospetto 3.2 - La classificazione della condizione di precarietà lavorativa**

Tipologia lavorativa (di tipo 1)	Tipologia lavorativa (di tipo 2)	Altre caratteristiche osservate	Condizione di precarietà (1)	Condizione di precarietà (2)
Dipendente a tempo indeterminato	Dipendente a tempo indeterminato - Full Time	Presenza di CIG	Lavoro stabile	Lavoro stabile (non sottoposto a CIG)
	Dipendente a tempo indeterminato - Part Time	Presenza di CIG		Lavoro stabile sottoposto a CIG
Autonomo	Autonomo	-		
Altro lavoro dipendente	Altro lavoro dipendente	Presenza di CIG	Altro lavoro dipendente	Altro lavoro dipendente
Dipendente a tempo determinato	Dipendente a tempo determinato	Presenza di CIG Mono/pluri committenza	Lavoro precario	Lavoro precario - Dipendente
Collaboratore	Collaboratore	Mono/pluri committenza		Lavoro precario - Collaboratore
Lavoratore domestico	Lavoratore domestico	Mono/pluri committenza		Lavoro precario - Domestico

### 3.2.2 L'identificazione del lavoratore precario

Osservati i contenuti informativi dell'output, è opportuno richiamare le modalità con le quali è stato individuato il lavoratore precario, a partire dagli aspetti definatori già esposti in precedenza secondo i quali è classificato precario il lavoratore che nell'arco di tempo considerato svolge esclusivamente lavori definiti precari.

Numerose sono le possibilità di classificare le tipologie contrattuali individuate nelle fonti studio utilizzate: dal punto di vista normativo si è fatto riferimento ai contratti collettivi nazionali, anche di secondo livello; dal punto di vista operativo, sono state analizzate le figure professionali; inoltre le fonti amministrative consentono di associare l'individuo alla propria cassa previdenziale. L'individuazione del lavoro precario è avvenuta selezionando categorie contrattuali che tengono conto congiuntamente di due aspetti:

- il termine contrattuale: distinzione tra lavoro standard e non standard;
- la professionalità: specializzazione ed esperienza specifica necessaria allo svolgimento dell'attività lavorativa.

Particolare attenzione è stata rivolta alla classificazione degli occupati presenti nell'archivio INPS dei Lavoratori Domestici; con riferimento alle peculiarità previste dal loro contratto collettivo nazionale in relazione alla risoluzione del rapporto di lavoro,<sup>39</sup> ai lavoratori domestici è stata attribuita la condizione "precario" di lavoro, anche se formalmente essi possono essere assunti con contratti di lavoro a tempo indeterminato.

Il prospetto seguente mostra come, attraverso la ricognizione di categorie contrattuali atipiche in vigore per le annualità 2010 e 2011, sia stato individuato l'insieme di quelle connesse al lavoro precario.

<sup>39</sup> All'art. 39 "Risoluzione del rapporto di lavoro e preavviso", co. 1, del CCNL sulla Disciplina del Rapporto di Lavoro Domestico (rinnovo triennio 2013-2016), si legge: "Il rapporto di lavoro può essere risolto da ciascuna delle parti con l'osservanza dei seguenti termini di preavviso. (1) Per i rapporti non inferiori a 25 ore settimanali, fino a 5 anni di anzianità presso lo stesso datore di lavoro: 15 giorni di calendario; oltre i 5 anni di anzianità presso lo stesso datore di lavoro: 30 giorni di calendario. I suddetti termini saranno ridotti del 50% nel caso di dimissioni da parte del lavoratore. (2) Per i rapporti inferiori alle 25 ore settimanali, fino a 2 anni di anzianità presso lo stesso datore di lavoro: 8 giorni di calendario; oltre i 2 anni di anzianità presso lo stesso datore di lavoro: 15 giorni di calendario".

**Prospetto 3.3 - L'individuazione delle categorie contrattuali di lavoro precario**

Categoria di lavoro individuata	Categoria contrattuale / rapporto di lavoro definiti "precarì"	Tipologia di lavoro classificata nell'output
Lavoro dipendente	Lavoro a tempo determinato (Full time e Part time), inclusi contratti di somministrazione, esclusi lavoratori domestici	Dipendente a tempo determinato
	Contratto di Formazione e Lavoro - CFL	Dipendente a tempo determinato
	Apprendistato	Dipendente a tempo determinato
	Contratto d'inserimento	Dipendente a tempo determinato
	Lavoro domestico a tempo determinato ed indeterminato	Lavoro domestico
Lavoro parasubordinato (fonte: Gestione Separata INPS)	Collaboratore di giornali	Collaboratore
	Dottorato di ricerca, assegno, borsa di studio	Collaboratore
	Co.co.co. (con contratto a progetto)	Collaboratore
	Venditore porta a porta	Collaboratore
	Collaboratori occasionali (mini Co.co.co.)	Collaboratore
	Rapporti occasionali autonomi	Collaboratore
	Co.co.co. presso la PA	Collaboratore
	Rapporti Co.co.co. prorogarti	Collaboratore
	Associati in partecipazione	Collaboratore
	Medici in formazione specialistica	Collaboratore
	Lavoro occasionale di tipo accessorio – Voucher	Collaboratore
Lavoro autonomo	P.IVA con mono committenza (risultato atteso con le attività dell'anno 2014)	Collaboratore (classificato Autonomo in fase sperimentale)

**3.2.3 L'integrazione delle fonti e la qualità dell'output**

La costruzione del data base di dati integrati per la sperimentazione ha tenuto conto di numerosi aspetti connessi al contenuto e alla qualità delle informazioni incluse nelle fonti amministrative utilizzate. A valle delle analisi di qualità dei microdati, si sono rese necessarie attività di loro integrazione, omogeneizzazione e trasformazione. In particolare nei passi di integrazione delle fonti sono state risolte le seguenti questioni:

- controllo e correzione delle incoerenze presenti nella stessa fonte nel medesimo periodo di riferimento;
- eventuali duplicazioni di informazione (ad es. sovrapposizioni di attività lavorative svolte contemporaneamente presso lo stesso datore, ovvero attività incompatibili svolte simultaneamente);
- variazioni delle informazioni amministrative contenute nella stessa fonte in periodi differenti, a seguito di cambiamenti nei sistemi di raccolta dati, nonché alla luce di aggiornamenti normativi, soprattutto in tema di mercato del lavoro;
- ricostruzione della posizione lavorativa mensile (non tiene conto di attività lavorative consecutive, per via dell'indisponibilità informativa di alcune fonti: ad esempio i Lavoratori Autonomi in Agricoltura sono considerati occupati per l'intero anno di dichiarazione di versamento contributivo).

Il livello di qualità delle informazioni presenti nel SIM è piuttosto differenziato. Se da un lato la fonte ASIA Occupazione era stata già sottoposta ad un controllo di qualità e coerenza dei dati piuttosto avanzato, le restanti fonti utilizzate sono state trattate in via sperimentale. In sintesi, è stata rilevata ed esaminata la presenza dei seguenti aspetti:

- codici fiscali, degli occupati e dei datori di lavoro, mancanti o errati;
- la fornitura della fonte Ex INPDAP è stata aggiornata di recente relativamente agli anni sui quali è stata effettuata la sperimentazione, e richiederà uno studio specifico previsto con le attività dell'anno 2014;

- variabili per le quali sono stati osservati valori apparentemente anomali che richiedono l'approfondimento (l'incoerenza di una condizione contrattuale di uno stesso individuo rilevata in un mese rispetto a quella rilevata nei restanti mesi dell'anno si può individuare, ed eventualmente correggere, solo analizzando l'intero periodo).

Una soddisfacente e sintetica evidenza della qualità dei risultati della sperimentazione è rappresentato dal contenuto della seguente tavola nella quale sono posti a confronto diretto i risultati del presente sotto-progetto e i dati delle Forze di Lavoro, principale indagine sul mercato del lavoro italiano condotta dall'Istat.<sup>40</sup>

**Tavola 3.1 - Confronto risultati principali del sotto-progetto Precarietà lavorativa e dati di fonte RCFL. (Valori assoluti in migliaia e val. %)**

PERIODO DI RIFERIMENTO	ARCHIMEDE – Precarietà lavorativa					Forze di Lavoro – Dati mensili (non destagionalizzati)				
	Maschi	%	Femmine	%	Totale	Maschi	%	Femmine	%	Totale
Ottobre 2010	13.039	57,8%	9.528	42,2%	22.568	13.702	59,8%	9.205	40,2%	22.907
Ottobre 2011	12.895	57,5%	9.521	42,5%	22.417	13.668	59,1%	9.441	40,9%	23.109

PERIODO DI RIFERIMENTO	ARCHIMEDE – Precarietà lavorativa					Forze di Lavoro – Dati trimestrali (IV Trim. non destagionalizzato)				
	Dipendenti	%	Indip.ti	%	Totale	Dipendenti	%	Indip.ti	%	Totale
Ottobre 2010	16.866	74,7%	5.702	25,3%	22.568	17.290	75,4%	5.645	24,6%	22.935
Ottobre 2011	16.768	74,8%	5.649	25,2%	22.417	17.385	75,7%	5.568	24,3%	22.953

### 3.3 Le potenzialità informative della base dei dati

I prodotti sperimentali sono stati elaborati al fine di mettere in evidenza le potenzialità delle fonti integrate riguardo all'esplorazione statistica del fenomeno della "Precarietà lavorativa". Nello specifico le tavole seguenti sono state realizzate mediante l'uso dell'Output di tipo 2 che rappresenta l'universo degli occupati nei mesi di ottobre 2010 e 2011 e sono corredate di informazioni relative alla loro condizione lavorativa principale e ad altre eventuali attività di lavoro svolte. Le elaborazioni riguardano: analisi cross section; analisi dei flussi di entrata/uscita dalla condizione di occupato; analisi longitudinali e prime analisi sui redditi.

#### Le analisi di tipo *cross section*

L'analisi del numero di occupati classificati per tipologia di lavoro e per sesso (Tavola 3.2) mostra come la quota di donne sia maggiore tra i lavoratori part time e i domestici, mentre, al contrario, i lavoratori full time e autonomi sono in prevalenza uomini. Considerando la condizione di precarietà (Tavola 3.3) si presenta una inversione di tendenza rispetto alla composizione dell'occupazione stabile per sesso: sono le donne quelle che possiedono le maggioranze di contratti precari (57% sia nel 2010, sia nel 2011). L'età media (Tavola 3.4) degli occupati stabili è sensibilmente più elevata di quella degli occupati precari (43 anni a fronte di 38 anni). Con riferimento alla mono/pluricommitenza (Tavola 3.5), circa il 9% dei lavoratori precari ha prestato la propria attività per più di un committente contro circa il 4% dei lavoratori stabili; mentre tra i lavoratori precari pluricommittenti sono più rappresentati i dipendenti (circa il 10% nel 2010 e oltre l'8% nel 2011) rispetto ai collaboratori (poco più del 3% in entrambi gli anni).

<sup>40</sup> Per chiarezza, al netto dei periodi di riferimento confrontati, vanno elencati sinteticamente alcuni punti utili ad ipotizzare la natura delle differenze tra i dati riportati in tabella: (1. Aspetti metodologici) la Rilevazione Continua sulle Forze di Lavoro (RCFL) è una indagine campionaria, mentre i dati utilizzati nel progetto sulla "Precarietà lavorativa" sono universali e di natura amministrativa, e pertanto non si è in grado di osservare il fenomeno del lavoro "sommerso"; le informazioni sulle forze di lavoro sono rilevate da percezione soggettiva - occupazione e relative caratteristiche - mentre i dati amministrativi hanno validità fiscale/contributiva; (2. Popolazione di riferimento e rilevata) i dati amministrativi sono relativi alle imprese residenti sul territorio italiano, mentre le forze di lavoro sono relative alle famiglie residenti nei comuni italiani; riguardo la fonte ASIA Occupazione non sono stati considerati circa 40.000 lavoratori dipendenti l'anno, in quanto operativamente imputati a causa dell'indisponibilità del loro microdato; la fonte Posizioni Assicurative Ex INPDAP non ricomprende circa 120.000 occupati militari del Ministero della Difesa - Esercito, in quanto le loro DMA (Denunce Mensili Analitiche) confluiranno nella fornitura presumibilmente a partire dall'anno di rilevazione 2012.



**Tavola 3.2 - Occupati per tipologia di lavoro e sesso. (Valori assoluti in migliaia e val. %)**

ANNO 2010 TIPOLOGIA	Maschi		Femmine		Totale
	V. a.	%	V. a.	%	V. a.
Dip.te T. Ind.to - Full time	7.551	63,3%	4.377	36,7%	11.928
Dip.te T. Ind.to - Part time	542	23,5%	1.765	76,5%	2.308
<b>Dip.te T. Ind.to - Totale</b>	<b>8.093</b>	<b>56,9%</b>	<b>6.143</b>	<b>43,1%</b>	<b>14.236</b>
Autonomo	3.646	69,5%	1.600	30,5%	5.246
Altro dip.te	17	21,8%	61	78,2%	78
Dip.te T. Det.to	973	52,1%	894	47,9%	1.866
Collaboratore	215	47,2%	241	52,8%	456
Lav.re Domestico	95	13,9%	590	86,1%	685
<b>Totale</b>	<b>13.039</b>	<b>57,8%</b>	<b>9.528</b>	<b>42,2%</b>	<b>22.568</b>

ANNO 2011 TIPOLOGIA	Maschi		Femmine		Totale
	V. a.	%	V. a.	%	V. a.
Dip.te T. Ind.to - Full time	7.428	63,0%	4.365	37,0%	11.793
Dip.te T. Ind.to - Part time	564	23,9%	1.797	76,1%	2.361
<b>Dip.te T. Ind.to - Totale</b>	<b>7.992</b>	<b>56,5%</b>	<b>6.162</b>	<b>43,5%</b>	<b>14.155</b>
Autonomo	3.622	69,4%	1.599	30,6%	5.221
Altro dip.te	18	21,8%	63	78,2%	81
Dip.te T. Det.to	997	53,3%	871	46,7%	1.868
Collaboratore	193	45,1%	235	54,9%	428
Lav.re Domestico	74	11,1%	590	88,9%	664
<b>Totale</b>	<b>12.895</b>	<b>57,5%</b>	<b>9.521</b>	<b>42,5%</b>	<b>22.417</b>

**Tavola 3.3 - Occupati per condizione di precarietà e sesso. (Valori assoluti in migliaia e val. %)**

ANNO 2010 LAVORATORE	Maschi		Femmine		Non classificato		Totale
	V. a.	%	V. a.	%	V. a.	%	V. a.
STABILE non CIG	11.321	59,8%	7.599	40,2%	0	0,0%	18.920
STABILE in CIG	418	74,3%	144	25,7%	0	0,0%	562
<b>STABILE - Totale</b>	<b>11.739</b>	<b>60,3%</b>	<b>7.743</b>	<b>39,7%</b>	<b>0</b>	<b>0,0%</b>	<b>19.482</b>
Altro dip.te	17	21,8%	61	78,2%	0	0,0%	78
PRECARIO dip.te	973	52,1%	894	47,9%	0	0,0%	1.866
PRECARIO collab.re	215	47,2%	241	52,8%	0	0,0%	456
PRECARIO domestico	95	13,9%	590	86,1%	0	0,0%	685
<b>PRECARIO - Totale</b>	<b>1.283</b>	<b>42,7%</b>	<b>1.725</b>	<b>57,3%</b>	<b>0</b>	<b>0,0%</b>	<b>3.008</b>
<b>TOTALE</b>	<b>13.039</b>	<b>57,8%</b>	<b>9.528</b>	<b>42,2%</b>	<b>0</b>	<b>0,0%</b>	<b>22.568</b>

ANNO 2011 LAVORATORE	Maschi		Femmine		Non classificato		Totale
	V. a.	%	V. a.	%	V. a.	%	V. a.
STABILE non CIG	11.269	59,6%	7.624	40,4%	0	0,0%	18.893
STABILE in CIG	345	71,5%	138	28,5%	0	0,0%	483
<b>STABILE - Totale</b>	<b>11.614</b>	<b>59,9%</b>	<b>7.761</b>	<b>40,1%</b>	<b>0</b>	<b>0,0%</b>	<b>19.375</b>
Altro dip.te	18	21,8%	63	78,2%	0	0,0%	81
PRECARIO dip.te	997	53,3%	871	46,7%	0	0,0%	1.868
PRECARIO collab.re	193	45,1%	235	54,9%	0	0,0%	428
PRECARIO domestico	74	11,1%	590	88,9%	0	0,0%	664
<b>PRECARIO - Totale</b>	<b>1.264</b>	<b>42,7%</b>	<b>1.697</b>	<b>57,3%</b>	<b>0</b>	<b>0,0%</b>	<b>2.960</b>
<b>TOTALE</b>	<b>12.895</b>	<b>57,5%</b>	<b>9.521</b>	<b>42,5%</b>	<b>0</b>	<b>0,0%</b>	<b>22.417</b>

**Tavola 3.4 - Occupati per condizione di precarietà ed età media. (Valori assoluti in migliaia e val. %)**

LAVORATORE	Anno 2010				Anno 2011			
	Totale		Età media	Non classificato	Totale		Età media	Non classificato
	V. a.	%			V. a.	%		
STABILE non CIG	18.920	83,8%	43,1	3	18.893	84,3%	43,5	2
STABILE in CIG	562	2,5%	42,1	0	483	2,2%	42,7	0
<b>STABILE - Totale</b>	<b>19.482</b>	<b>86,3%</b>	<b>43,0</b>	<b>3</b>	<b>19.375</b>	<b>86,4%</b>	<b>43,4</b>	<b>2</b>
Altro dip.te	78	0,3%	43,5	0	81	0,4%	43,7	0
PRECARIO dip.te	1.866	8,3%	36,4	1	1.868	8,3%	36,3	1
PRECARIO collab.re	456	2,0%	38,5	0	428	1,9%	37,6	0
PRECARIO domestico	685	3,0%	42,9	0	664	3,0%	44,0	0
<b>PRECARIO - Totale</b>	<b>3.008</b>	<b>13,3%</b>	<b>38,0</b>	<b>1</b>	<b>2.960</b>	<b>13,2%</b>	<b>38,2</b>	<b>1</b>
<b>TOTALE</b>	<b>22.568</b>	<b>100,0%</b>	<b>42,4</b>	<b>4</b>	<b>22.417</b>	<b>100,0%</b>	<b>42,8</b>	<b>3</b>

**Tavola 3.5 - Occupati per condizione di precarietà e numero di committenti. (Valori assoluti in migliaia e val. %)**

ANNO 2010 LAVORATORE	Un committente		Due o più committenti		Non classificato		Totale
	V. a.	%	V. a.	%	V. a.	%	V. a.
STABILE non CIG	18.041	95,4%	852	4,5%	27	0,1%	18.920
STABILE in CIG	547	97,3%	15	2,6%	0	0,1%	562
<b>STABILE - Totale</b>	<b>18.588</b>	<b>95,4%</b>	<b>866</b>	<b>4,4%</b>	<b>27</b>	<b>0,1%</b>	<b>19.482</b>
Altro dip.te	70	90,0%	3	3,5%	5	6,5%	78
PRECARIO dip.te	1.663	89,1%	200	10,7%	3	0,2%	1.866
PRECARIO collab.re	437	95,9%	16	3,4%	3	0,6%	456
PRECARIO domestico	620	90,5%	65	9,5%	0	0,0%	685
<b>PRECARIO - Totale</b>	<b>2.721</b>	<b>90,5%</b>	<b>281</b>	<b>9,3%</b>	<b>6</b>	<b>0,2%</b>	<b>3.008</b>
<b>TOTALE</b>	<b>21.379</b>	<b>94,7%</b>	<b>1.150</b>	<b>5,1%</b>	<b>39</b>	<b>0,2%</b>	<b>22.568</b>
ANNO 2011 LAVORATORE	Un committente		Due o più committenti		Non classificato		Totale
	V. a.	%	V. a.	%	V. a.	%	V. a.
STABILE non CIG	18.047	95,5%	827	4,4%	19	0,1%	18.893
STABILE in CIG	470	97,3%	13	2,7%	0	0,0%	483
<b>STABILE - Totale</b>	<b>18.517</b>	<b>95,6%</b>	<b>840</b>	<b>4,3%</b>	<b>19</b>	<b>0,1%</b>	<b>19.375</b>
Altro dip.te	74	90,7%	3	4,2%	4	5,1%	81
PRECARIO dip.te	1.698	90,9%	166	8,9%	3	0,2%	1.868
PRECARIO collab.re	413	96,4%	13	3,1%	2	0,5%	428
PRECARIO domestico	594	89,4%	70	10,6%	0	0,0%	664
<b>PRECARIO - Totale</b>	<b>2.705</b>	<b>91,4%</b>	<b>250</b>	<b>8,4%</b>	<b>5</b>	<b>0,2%</b>	<b>2.960</b>
<b>TOTALE</b>	<b>21.296</b>	<b>95,0%</b>	<b>1.093</b>	<b>4,9%</b>	<b>28</b>	<b>0,1%</b>	<b>22.417</b>

### Le analisi di flusso entrate/uscite da/verso condizione lavorativa

Con riferimento ai flussi per condizione lavorativa che si presentano fra l'Ottobre 2010 e l'Ottobre 2011 (Tavole da 3.6 a 3.9), la quota maggiore di uscite si verifica tra gli occupati stabili (oltre il 68% a fronte del 31% tra i precari); sempre tra gli usciti stabili, la quota maggiore si registra tra gli uomini mentre tra gli usciti precari prevalgono le donne. L'età media degli usciti tra gli occupati stabili è sensibilmente più elevata di quella degli usciti tra gli occupati precari. Con riferimento alle entrate, la quota maggiore si verifica tra gli occupati stabili (circa 58% a fronte del 42% tra i precari), mentre con riferimento al genere, tra gli entrati stabili la quota maggiore si registra tra gli uomini mentre tra gli entrati precari prevalgono le donne. L'età media degli entrati tra gli occupati stabili è più elevata di quella degli entrati tra gli occupati precari.

**Tavola 3.6 - Occupati nel 2010 usciti dall'occupazione nel 2011 per condizione di precarietà e sesso. (Valori assoluti e %)**

LAVORATORE	MASCHI		FEMMINE		Totale	
	V.a.	%	V.a.	%	V.a.	%
STABILE non CIG	911	70,1%	602	59,6%	1.514	65,5%
STABILE in CIG	54	4,1%	20	2,0%	74	3,2%
<b>STABILE - Totale</b>	<b>965</b>	<b>74,2%</b>	<b>622</b>	<b>61,6%</b>	<b>1.588</b>	<b>68,7%</b>
Altro dip.te	2	0,2%	4	0,4%	7	0,3%
PRECARIO dip.te	235	18,1%	198	19,6%	433	18,7%
PRECARIO collab.re	67	5,2%	69	6,8%	136	5,9%
PRECARIO domestico	30	2,3%	117	11,6%	147	6,4%
<b>PRECARIO - Totale</b>	<b>333</b>	<b>25,6%</b>	<b>384</b>	<b>38,0%</b>	<b>717</b>	<b>31,0%</b>
<b>TOTALE</b>	<b>1.300</b>	<b>100,0%</b>	<b>1.011</b>	<b>100,0%</b>	<b>2.311</b>	<b>100,0%</b>

**Tavola 3.7 - Occupati nel 2010 usciti dall'occupazione nel 2011 per condizione di precarietà ed età media. (Valori assoluti e %)**

LAVORATORE	Totale		Età media	Non classificato
	V. a.	%		
STABILE non CIG	1.514	65,5%	43,7	1
STABILE in CIG	74	3,2%	44,1	0
<b>STABILE - Totale</b>	<b>1.588</b>	<b>68,7%</b>	<b>43,8</b>	<b>1</b>
Altro dip.te	7	0,3%	44,2	0
PRECARIO dip.te	433	18,7%	36	1
PRECARIO collab.re	136	5,9%	40,7	0
PRECARIO domestico	147	6,4%	41,2	0
<b>PRECARIO - Totale</b>	<b>717</b>	<b>31,0%</b>	<b>38</b>	<b>1</b>
<b>TOTALE</b>	<b>2.311</b>	<b>100,0%</b>	<b>42</b>	<b>2</b>

**Tavola 3.8 - Inoccupati nel 2010 entrati nell'occupazione nel 2011 per condizione di precarietà e sesso. (Valori assoluti e %)**

LAVORATORE	Maschi		Femmine		Totale	
	V. a.	%	V. a.	%	V. a.	%
STABILE non CIG	731	63,2%	501	49,9%	1.232	57,0%
STABILE in CIG	10	0,9%	3	0,3%	13	0,6%
<b>STABILE - Totale</b>	<b>741</b>	<b>64,1%</b>	<b>504</b>	<b>50,2%</b>	<b>1.245</b>	<b>57,6%</b>
Altro dip.te	3	0,2%	6	0,6%	8	0,4%
PRECARIO dip.te	330	28,5%	289	28,8%	619	28,7%
PRECARIO collab.re	61	5,3%	82	8,2%	143	6,6%
PRECARIO domestico	22	1,9%	123	12,3%	145	6,7%
<b>PRECARIO - Totale</b>	<b>413</b>	<b>35,7%</b>	<b>495</b>	<b>49,3%</b>	<b>907</b>	<b>42,0%</b>
<b>TOTALE</b>	<b>1.156</b>	<b>100,0%</b>	<b>1.004</b>	<b>100,0%</b>	<b>2.160</b>	<b>100,0%</b>

**Tavola 3.9 - Inoccupati nel 2010 entrati nell'occupazione nel 2011 per condizione di precarietà ed età media. (Valori assoluti e %)**

LAVORATORE	Totale		Età media	Non classificato
	V. a.	%		
STABILE non CIG	1.232	57,0%	37,1	0
STABILE in CIG	13	0,6%	39,7	0
<b>STABILE - Totale</b>	<b>1.245</b>	<b>57,6%</b>	<b>37,1</b>	<b>0</b>
Altro dip.te	8	0,4%	39,8	0
PRECARIO dip.te	619	28,7%	33,7	1
PRECARIO collab.re	143	6,6%	33,3	0
PRECARIO domestico	145	6,7%	41,2	0
<b>PRECARIO - Totale</b>	<b>907</b>	<b>42,0%</b>	<b>34,8</b>	<b>1</b>
<b>TOTALE</b>	<b>2.160</b>	<b>100,0%</b>	<b>36,2</b>	<b>1</b>

### Le analisi longitudinali

L'analisi longitudinale è stata condotta sugli individui che risultano occupati in entrambi i periodi di riferimento (panel di circa 20 milioni e 256 mila occupati). La matrice di transizione 2010-2011 degli occupati per tipologia lavorativa, viene presentata nella tavola 3.11, mentre quella per condizione di precarietà nella tavola 3.12. Ad un anno di distanza, la quasi totalità di dipendenti a tempo indeterminato (97%) si colloca nella medesima tipologia lavorativa, mentre circa il 34% dei dipendenti a tempo determinato è transitato verso una condizione contrattuale a tempo indeterminato. Con riferimento ai collaboratori, il 74% permane nella stessa condizione, oltre il 21% passano ad avere un contratto di lavoro dipendente (10,5% a tempo indeterminato) e il 5% passa nella condizione di lavoratore autonomo. Il maggior numero di transizioni da precario a stabile e viceversa (Tavole 3.12 e 3.13) si registra nella classe di età 25-34 anni e in quella 35-44 anni, mentre coloro che erano precari nel 2010 e permangono in questa condizione nel 2011 si collocano nelle fasce d'età comprese tra 25 e 54 anni e sono prevalentemente di genere femminile. Infine si evidenzia come ad un anno di distanza il 26,4% delle transizioni precario-stabile avviene nell'ambito dello stesso datore di lavoro mentre oltre il 75% delle permanenze nella condizione di precarietà coinvolgono lo stesso datore di lavoro.

**Tavola 3.10 - Occupati per tipologia lavorativa. (Valori assoluti in migliaia e val. %)**

		Anno 2011												
Anno 2010	Dip.te T. Ind.to	% riga	Autonomo	% riga	Altro dip.te	% riga	Dip.te T. Det.to	% riga	Collaboratore	% riga	Lav.re Domestico	% riga	Totale	% Totale
Dip.te T. Ind.to	12.670	97,0%	90	0,7%	6	0,0%	266	2,0%	20	0,2%	7	0,1%	13.059	64,5%
Autonomo	69	1,4%	4.720	97,6%	1	0,0%	33	0,7%	12	0,3%	1	0,0%	4.836	23,9%
Altro dip.te	6	9,0%	0	0,6%	60	84,6%	4	5,3%	0	0,2%	0	0,3%	71	0,4%
Dip.te T. Det.to	488	34,1%	19	1,3%	6	0,4%	901	62,9%	14	1,0%	5	0,4%	1.433	7,1%
Collaboratore	34	10,5%	16	4,9%	0	0,1%	32	9,9%	238	74,2%	1	0,4%	320	1,6%
Lav.re Domestico	17	3,2%	2	0,4%	0	0,0%	13	2,4%	1	0,2%	505	93,8%	538	2,7%
<b>Totale</b>	<b>13.284</b>	<b>65,6%</b>	<b>4.847</b>	<b>23,9%</b>	<b>73</b>	<b>0,4%</b>	<b>1.249</b>	<b>6,2%</b>	<b>285</b>	<b>1,4%</b>	<b>519</b>	<b>2,6%</b>	<b>20.256</b>	<b>100,0%</b>

**Tavola 3.11 - Occupati per condizione di precarietà. (Valori assoluti in migliaia e val. %)**

		Anno 2011																
Anno 2010	STABILE non CIG	% riga	STABILE in CIG	% riga	STABILE - LE - Totale	% riga	Altro dip.te	% riga	PRECA- RIO dip.te	% riga	PRECA- RIO col- lab.re	% riga	PRECA- RIO do- mestico	% riga	PRECA- RIO - Totale	% riga	Totale	% Totale
STABILE non CIG	16.863	96,9%	226	1,3%	17.089	98,2%	6	0,0%	272	1,6%	31	0,2%	8	0,0%	311	1,8%	17.406	85,9%
STABILE in CIG	227	46,5%	233	47,6%	460	94,1%	1	0,1%	27	5,6%	1	0,1%	0	0,0%	28	5,7%	488	2,4%
<b>STABILE - Totale</b>	<b>17.090</b>	<b>95,5%</b>	<b>459</b>	<b>2,6%</b>	<b>17.548</b>	<b>98,1%</b>	<b>7</b>	<b>0,0%</b>	<b>299</b>	<b>1,7%</b>	<b>32</b>	<b>0,2%</b>	<b>8</b>	<b>0,0%</b>	<b>339</b>	<b>1,9%</b>	<b>17.894</b>	<b>88,3%</b>
Altro dip.te	7	9,4%	0	0,2%	7	9,9%	60	84,6%	4	5,3%	0	0,2%	0	0,3%	4	5,6%	71	0,4%
PRECARIO dip.te	496	34,6%	11	0,8%	507	35,4%	6	0,4%	901	62,9%	14	1,0%	5	0,4%	920	64,2%	1.433	7,1%
PRECARIO col- lab.re	49	15,3%	0	0,1%	49	15,3%	0	0,1%	32	9,9%	238	74,2%	1	0,4%	270	84,4%	320	1,6%
PRECARIO do- mestico	19	3,6%	0	0,0%	19	3,6%	0	0,0%	13	2,4%	1	0,2%	505	93,8%	519	96,5%	538	2,7%
<b>PRECARIO - Totale</b>	<b>564</b>	<b>24,6%</b>	<b>11</b>	<b>0,5%</b>	<b>575</b>	<b>25,1%</b>	<b>6</b>	<b>0,3%</b>	<b>946</b>	<b>41,3%</b>	<b>253</b>	<b>11,0%</b>	<b>511</b>	<b>22,3%</b>	<b>1.709</b>	<b>74,6%</b>	<b>2.291</b>	<b>11,3%</b>
<b>Totale</b>	<b>17.661</b>	<b>87,2%</b>	<b>470</b>	<b>2,3%</b>	<b>18.131</b>	<b>89,5%</b>	<b>73</b>	<b>0,4%</b>	<b>1.249</b>	<b>6,2%</b>	<b>285</b>	<b>1,4%</b>	<b>519</b>	<b>2,6%</b>	<b>2.053</b>	<b>10,1%</b>	<b>20.256</b>	<b>100,0%</b>

**Tavola 3.12 - Transizioni di condizione precarietà per classi di età e sesso. (Valori assoluti in migliaia e val. %)**

<b>Occupati in transizione PRECARIO 2010 - STABILE 2011</b>						
CLASSE DI ETÀ	Maschi		Femmine		Totale	
	V. a.	%	V. a.	%	V. a.	
15-24	47	58,0%	34	42,0%	82	
25-34	112	53,9%	96	46,1%	208	
35-44	84	51,9%	78	48,1%	162	
45-54	48	50,9%	47	49,1%	95	
55-64	16	62,5%	10	37,5%	26	
65+	2	81,1%	0	18,9%	2	
Non classificato	0	45,5%	0	54,5%	0	
<b>Totale</b>	<b>310</b>	<b>53,9%</b>	<b>265</b>	<b>46,1%</b>	<b>575</b>	
<b>Di cui con stesso identificativo di datore di lavoro</b>					<b>152 (26,4%)</b>	

<b>Occupati in transizione STABILE 2010 - PRECARIO 2011</b>						
CLASSE DI ETÀ	Maschi		Femmine		Totale	
	V. a.	%	V. a.	%	V. a.	
15-24	27	61,2%	17	38,8%	45	
25-34	65	58,5%	47	41,5%	112	
35-44	62	61,7%	38	38,3%	100	
45-54	38	65,4%	20	34,6%	58	
55-64	17	75,3%	5	24,7%	22	
65+	2	83,1%	0	16,9%	2	
Non classificato	0	75,0%	0	25,0%	0	
<b>Totale</b>	<b>211</b>	<b>62,2%</b>	<b>128</b>	<b>37,8%</b>	<b>340</b>	
<b>Di cui con stesso identificativo di datore di lavoro</b>					<b>28 (8,2%)</b>	

<b>Occupati in transizione PRECARIO 2010 - PRECARIO 2011</b>						
CLASSE DI ETÀ	Maschi		Femmine		Totale	
	V. a.	%	V. a.	%	V. a.	
15-24	79	50,4%	78	49,6%	157	
25-34	210	40,5%	309	59,5%	519	
35-44	165	34,3%	315	65,7%	480	
45-54	109	29,6%	260	70,4%	369	
55-64	61	38,0%	100	62,0%	162	
65+	14	62,1%	9	37,9%	23	
Non classificato	0	70,0%	0	24,7%	0	
<b>Totale</b>	<b>639</b>	<b>37,4%</b>	<b>1.071</b>	<b>62,6%</b>	<b>1.709</b>	
<b>Di cui con stesso identificativo di datore di lavoro</b>					<b>1.284 (75,1%)</b>	

### Le integrazioni con altre informazioni

Nel suo complesso, il progetto ARCHIMEDE comprende più attività in grado di esplorare diversi fenomeni mediante l'analisi dei contenuti informativi presenti in Istat. Tale prerogativa ha dato l'opportunità alle risorse umane impiegate nel progetto di lavorare in sinergia: riguardo ognuno dei sotto prodotti è stato possibile sfruttare in parte le attività specifiche svolte relativamente agli

altri sotto prodotti. Nello specifico, gli output del prodotto sulla “Precarietà lavorativa” sono stati integrati con le informazioni reddituali esaminate durante lo svolgimento del lavoro su “Precarietà familiare”.<sup>41</sup> La tavola 3.13 mostra la distribuzione del reddito mediano, da lavoro e complessivo, per le differenti tipologie di occupati dipendenti ed assimilati. Risulta evidente come il *rank* definito nella condizione sia confermato dal reddito mediano. Pur con questa evidenza, la quota di dipendenti precari con reddito superiore o uguale a quello mediano degli occupati stabili è del 13,5% (Tavola 3.14). Al contrario sono pari al 9,1% gli occupati stabili, con contratto full-time, che possiedono un reddito inferiore o uguale a quello mediano degli occupati precari. Tali percentuali sono significativamente maggiori (rispettivamente il 18% e il 20,1%) se si considerano il sottoinsieme dei pluricommittenti.

**Tavola 3.13 - Analisi reddituale sugli occupati dipendenti e assimilati (espressi in migliaia) ad ottobre 2011 per tipologia lavorativa e condizione di precarietà**

CONDIZIONE	Totale occupati dipendenti e assimilati	Di cui con reddito positivo	Reddito da lavoro dipendente o assimilato (val. mediano in Euro)	Reddito complessivo (val. mediano in Euro)
Dip.te T. Ind.to - Full time	11.793	11.656	22.390	22.985
Dip.te T. Ind.to - Part time	2.361	2.319	10.818	11.466
<b>STABILE - Totale</b>	<b>14.154</b>	<b>13.975</b>	<b>20.764</b>	<b>21.352</b>
Altro dip.te	81	80	13.758	14.624
Dip.te T. Det.to	1.868	1.825	13.009	13.590
Collaboratore	428	348	8.317	10.515
Lav.re Domestico	664	149	7.447	7.981
<b>PRECARIO - Totale</b>	<b>2.960</b>	<b>2.323</b>	<b>11.705</b>	<b>12.558</b>
<b>TOTALE</b>	<b>17.195</b>	<b>16.377 (95,2%)</b>	<b>19.616</b>	<b>20.183</b>

<sup>41</sup> Tali informazioni per l'anno 2011 provengono dalla Banca Dati Redditali (BDR) del Ministero dell'Economia e delle Finanze (MEF). Per reddito da lavoro dipendente o assimilato si intende la somma dei redditi da lavoro dipendente e assimilato, inclusi gli straordinari; è una variabile ricalcolata dal MEF a partire dalle informazioni contenute nel Modello Unico (Sez. I, Quadro RC) per la dichiarazione dei redditi. La popolazione di riferimento comprende lavoratori dipendenti in senso stretto, soci di cooperative di produzione e lavoro, lavoratori socialmente utili, co.co.co., co.co.pro., sacerdoti e altri ministri di culto. Il reddito complessivo (al lordo delle deduzioni), anch'esso rideterminato dal MEF, include i redditi a tassazione ordinaria più la cedolare secca: è dato dalla somma dei singoli redditi indicati nei vari quadri del Modello Unico (dominicali, agrari, fabbricati, lavoro dipendente, lavoro autonomo, impresa in contabilità semplificata o ordinaria, imprese consorziate, partecipazione, plusvalenze di natura finanziaria, allevamento, redditi a tassazione separata con opzione tassazione ordinaria).

**Tavola 3.14 - Analisi reddituale sugli occupati dipendenti e collaboratori (espressi in migliaia) ad ottobre 2011 per tipologia lavorativa. Confronto tra redditi da lavoro dipendente e assimilato mediani per condizione di precarietà e mono-pluricommittenza (val. %)**

<b>Dipendenti e collaboratori TOTALI con reddito positivo</b>			
TIPOLOGIA	Occupati	Quota % di soggetti con reddito superiore o uguale a quello mediano degli STABILI	Quota % di soggetti con reddito inferiore a quello mediano dei PRECARI
Dip.te T. Ind.to - Full time	11.656	58,5%	9,1%
Dip.te T. Ind.to - Part time	2.319	7,1%	56,9%
<b>STABILE - Totale</b>	<b>13.975</b>	<b>50,0%</b>	<b>17,0%</b>
Altro dip.te	80	16,4%	33,7%
Dip.te T. Det.to	1.825	14,2%	44,6%
Collaboratore	348	15,1%	63,4%
Lav.re Domestico	149	0,5%	85,1%
<b>PRECARIO - Totale</b>	<b>2.323</b>	<b>13,5%</b>	<b>50,0%</b>
<b>TOTALE</b>	<b>16.377</b>	<b>44,7%</b>	<b>21,8%</b>

<b>Dipendenti e collaboratori MONOCOMMITTENTI con reddito positivo</b>			
TIPOLOGIA	Occupati	Quota % di soggetti con reddito superiore o uguale a quello mediano degli STABILI	Quota % di soggetti con reddito inferiore a quello mediano dei PRECARI
Dip.te T. Ind.to - Full time	11.373	58,6%	8,8%
Dip.te T. Ind.to - Part time	2.158	6,6%	57,5%
<b>STABILE - Totale</b>	<b>13.531</b>	<b>50,3%</b>	<b>16,5%</b>
Altro dip.te	73	17,1%	31,2%
Dip.te T. Det.to	1.657	13,7%	45,7%
Collaboratore	333	14,8%	63,8%
Lav.re Domestico	133	0,6%	85,3%
<b>PRECARIO - Totale</b>	<b>2.124</b>	<b>13,1%</b>	<b>51,1%</b>
<b>TOTALE</b>	<b>15.728</b>	<b>45,1%</b>	<b>21,3%</b>

<b>Dipendenti e collaboratori PLURICOMMITTENTI con reddito positivo</b>			
TIPOLOGIA	Occupati	Quota % di soggetti con reddito superiore o uguale a quello mediano degli STABILI	Quota % di soggetti con reddito inferiore a quello mediano dei PRECARI
Dip.te T. Ind.to - Full time	269	57,6%	20,1%
Dip.te T. Ind.to - Part time	158	13,5%	48,5%
<b>STABILE - Totale</b>	<b>427</b>	<b>41,3%</b>	<b>30,6%</b>
Altro dip.te	3	20,0%	36,1%
Dip.te T. Det.to	165	19,3%	32,7%
Collaboratore	13	23,3%	54,2%
Lav.re Domestico	16	0,2%	83,3%
<b>PRECARIO - Totale</b>	<b>194</b>	<b>18,0%</b>	<b>38,3%</b>
<b>TOTALE</b>	<b>624</b>	<b>18,0%</b>	<b>38,3%</b>

### 3.4 Sviluppi futuri

L' output sperimentale del progetto precarietà lavorativa è stato circoscritto all'analisi di aspetti strettamente connessi al contratto di lavoro e solo in parte a caratteri economici. Il lavoratore è stato pertanto classificato in condizione di precarietà alla luce di aspetti contrattuali che le prossime attività dovranno essere in grado di corredare con informazioni di carattere socio-economico.

Numerosi sono gli aspetti da approfondire al termine della prima fase sperimentale del progetto, sia rispetto all'ampliamento del periodo di riferimento osservato, sia riguardo i contenuti informa-



tivi esaminabili. La logica sarà quella di integrare le fonti demo-sociali ed economiche al fine di misurare il grado di precarietà del lavoratore, nonché il livello di sicurezza del lavoratore avente un contratto di lavoro stabile. Si potranno così calcolare specifici indicatori di precarietà (utilizzando informazioni sulle tipologie familiari, sul reddito e sui livelli di istruzione/formazione), ovvero misurare il livello di benessere dei lavoratori precari.

Seppur la classificazione dell'occupato, in quanto appartenente alle categorie di lavoratore precario o stabile, è alla base del progetto, lo studio delle transizioni lavorative impone di allargare lo sguardo alla condizione di non occupato. Si può infatti passare dalla condizione di occupato a quella di inoccupato e viceversa. Nel 2014 sarà analizzato il primo tipo di transizione usando i contenuti informativi esposti e descritti al paragrafo 3.1.3. Si cercherà dunque di determinare le destinazioni di stato di coloro che fuoriescono dalla condizione di occupato, distinguendo: pensionato; deceduto; emigrato; persona in cerca di lavoro; individuo appartenente alle non forze di lavoro. Tale risultato sarà raggiunto mediante l'integrazione degli output ottenuti durante la sperimentazione (opportunamente rivisti ed aggiornati) con altre fonti demo-sociali.

Tra le classi di occupati in relazione alle quali è stata demarcata la precarietà lavorativa compaiono le collaborazioni, incluse quelle autonome con P.IVA, che presentano circostanze di subordinazione. Attraverso l'esame delle categorie di rapporto di lavoro previste dalla fonte INPS – Gestione Separata Parasubordinati, è stato possibile individuare quelle afferenti a lavori definiti precari; lo studio dei lavoratori con P.IVA, ascrivibili alla categoria di lavoratori autonomi richiede un approfondimento ulteriore perché talvolta questo stato è caratterizzato da vincoli reali di subordinazione tali da aumentarne la percezione di precarietà. Infatti, sebbene il collaboratore autonomo con partita IVA debba svolgere la propria attività lavorativa in modo completamente indipendente, tale prestazione ha spesso costituito per i committenti un facile strumento di elusione delle norme di tutela del lavoro dipendente, mascherando veri e propri rapporti di lavoro subordinato.<sup>42</sup> Uno degli sviluppi futuri del presente progetto è proprio orientato all'individuazione dei lavoratori aventi P.IVA con mono committenza.

L'analisi longitudinale implementata durante la sperimentazione è stata circoscritta alla esclusiva osservazione dei mutamenti ad un certo periodo di distanza (arco di un anno: dati di *stock* ottobre 2010 – ottobre 2011). Le attività da svolgere durante l'anno 2014 verteranno invece sia sull'allargamento del campo di osservazione all'annualità 2012, sia all'approfondimento dell'analisi a periodi intermedi. Ricostruire i microdati sugli occupati per ogni mese dell'anno permetterà di esaminare questioni relative all'occupazione quali la mobilità del lavoro, la variazione delle retribuzioni e le disuguaglianze delle carriere lavorative degli individui e delle storie socio-economiche tra occupati precari e lavoratori stabili.

## 4. Il sotto-progetto sperimentale: la precarietà socio economica delle famiglie<sup>43</sup>

### 4.1 Le caratteristiche della sperimentazione

L'interesse per questa sperimentazione deriva dalla opportunità offerta dal Sistema Integrato di Microdati (SIM) dell'Istat di costruire una struttura informativa sulle famiglie mediante la quale analizzare vari aspetti delle loro condizioni socio-economiche. Più in particolare, si è inteso fare riferimento alla scomposizione del concetto di precarietà socio-economica in elementi osservabili e

<sup>42</sup> Uno degli obiettivi dichiarati della riforma del mercato del lavoro voluta dall'ex Ministro Fornero è proprio la lotta contro l'utilizzo improprio delle collaborazioni rese da titolari di partite IVA. L'art. 1, co. 26, della L. 92/2012, infatti, introducendo l'art. 69 bis del D.Lgs. 276/2003, secondo il quale l'attività professionale può essere trasformata in rapporto di co.co.co. (o addirittura in un contratto di lavoro dipendente a tempo indeterminato, se l'attività non è riconducibile ad un progetto specifico) quando sussistono almeno due dei seguenti presupposti: (1) collaborazione con durata complessivamente superiore ad otto mesi annui per due anni consecutivi; (2) guadagno, anche se fatturato a più soggetti riconducibili al medesimo centro d'imputazione di interessi, costituito da più dell'80% dei corrispettivi annui complessivamente percepiti dal collaboratore nell'arco di due anni solari consecutivi; (3) presenza di una postazione fissa di lavoro presso una delle sedi del committente.

<sup>43</sup> I §§ 4.1.1, 4.1.3, 4.2 e 4.4 sono a cura di D. Bonardo, i §§ 4.1.2, 4.1.4 e 4.3 sono a cura di S. Casacci.

misurabili a partire dai dati integrati provenienti da varie basi dati di origine amministrativa anche riferibili ad ambiti territoriali ristretti (comunali o sub-comunali).<sup>44</sup>

In letteratura i concetti connessi al tema della precarietà socio-economica sono quelli di “povertà”, “vulnerabilità sociale” ed “esclusione sociale”.<sup>45</sup> Per la Commissione Europea<sup>46</sup> la povertà non può essere definita semplicemente come mancanza di reddito e di risorse finanziarie, ma comprende anche la nozione di vulnerabilità e fattori quali il mancato accesso a un'alimentazione adeguata, all'istruzione e alla salute, alle risorse naturali e all'acqua potabile, alla terra, all'occupazione e al credito agevolato, all'informazione e alla partecipazione politica, ai servizi e alle infrastrutture. La vulnerabilità si configura come una condizione di rischio riconducibile ad una carenza di risorse in una o più dimensioni fondamentali dell'esistenza di un soggetto o di una famiglia. La ricerca sull'esclusione sociale include tra i suoi oggetti di studio anche individui che, pur non trovandosi in una situazione di povertà effettiva, vivono esperienze che possono innescare processi di impoverimento, soprattutto se ripetuti nel tempo.<sup>47</sup>

Il sotto-progetto ha l'obiettivo di produrre microdati riferiti alla popolazione di interesse e indicatori, a un dettaglio territoriale molto fine, con periodicità annuale e attraverso l'integrazione di contenuti informativi derivabili da fonti amministrative di natura socio-economica e demografica.

#### 4.1.1 Obiettivi

L'obiettivo principale è classificare tutte le famiglie residenti in un determinato territorio rispetto ad alcune dimensioni in modo da dedurre le loro condizioni di precarietà sociale ed economica e quelle degli individui che le compongono. Il punto di forza del progetto deriva dalla possibilità di combinare le informazioni che caratterizzano il concetto di precarietà a vari livelli di intensità.

Le dimensioni individuate, ritenute rappresentative per lo studio delle condizioni socio-economiche delle famiglie, sono sei:

- *Tipologia familiare*: struttura del nucleo familiare sulla base delle relazioni parentali dei membri che la compongono;
- *Reddito*: composizione e quantificazione della struttura del reddito complessivo di una famiglia;
- *Situazione abitativa*: caratteristiche dell'abitazione e tipologia del suo godimento;
- *Disagio*: insieme dei fattori che producono discriminazione ed esclusione sociale (ad es. la deprivazione economica o l'integrazione socioculturale);
- *Istruzione*: caratteristiche del percorso educativo e formativo dei componenti;
- *Lavoro*: descrizione delle caratteristiche dell'occupazione dei componenti del nucleo familiare.

Valorizzando i bacini informativi derivanti da fonti amministrative acquisite dall'Istat e integrate nel SIM è possibile produrre le seguenti basi informative di microdati:

1. una collezione di microdati avente come unità i singoli individui: questo primo output contiene le informazioni derivanti dalle variabili degli archivi amministrativi e rappresenta la base che consente la realizzazione del secondo output;
2. un dataset avente come unità le famiglie, costruito a partire da combinazioni e/o elaborazioni e/o sintesi di singole variabili del dataset individuale, in modo tale da poter individuare differenti situazioni di criticità delle famiglie.

<sup>44</sup> Attualmente, le stime derivanti dalle indagini campionarie forniscono stime per domini territoriali quali le regioni e le province. Ad esempio, le stime sul reddito delle famiglie derivanti dall'indagine campionaria Eu-Silc si fermano al livello regionale; l'indagine sulle Forze di Lavoro fornisce stime trimestrali a livello regionale e stime in media d'anno a livello provinciale.

<sup>45</sup> Sotto il profilo concettuale e definitorio l'inquadramento del fenomeno è stato effettuato a partire dai documenti redatti da: European Commission (2000); OECD (2001); Istat e Cnel (2013).

<sup>46</sup> European Commission (2000).

<sup>47</sup> Tsakloglou e Papadopoulos (2002).

#### 4.1.2 Unità statistiche, popolazione di riferimento e principali definizioni adottate

L'unità statistica primaria di questo studio è costituita dalla famiglia. Allo scopo di collezionare in modo organizzato i dati sui caratteri della popolazione di riferimento è tuttavia necessario partire dalla raccolta di informazioni sull'unità individuo, che si configura come unità di rilevazione e può essere un ulteriore oggetto di studio.

La popolazione di riferimento è rappresentata dalle famiglie (e dagli individui che le compongono) che vivono in un determinato ambito territoriale ad una data stabilita (possibilmente il 31 dicembre dell'anno di studio).

Un aspetto di rilievo del progetto riguarda la possibilità di identificare la struttura sociale della famiglia i cui membri sono uniti da legami di filiazione o adozione, di matrimonio, di affetto, indipendentemente dal fatto che condividano o meno lo stesso spazio domestico e che siano o meno registrati in anagrafe nella stessa famiglia. Dunque, l'intento dello studio è di fare riferimento non solo alla famiglia anagrafica e ai suoi eventuali nuclei, ma anche a grappoli di famiglie unite da legami di tipo economico-fiscale. Per raggiungere tale obiettivo è opportuno mettere a punto pertinenti metodologie di analisi che, sfruttando in modo integrato variabili contenute in diversi fonti, possano mettere in connessione persone residenti in luoghi diversi. Le attività finora realizzate mostrano una potenzialità ancora inespressa dei dati fiscali, che sarà certamente oggetto di ulteriori analisi. In questa prima fase della sperimentazione ci si è ancorati ai concetti già definiti in letteratura e nell'ambito della statistica ufficiale. Di seguito le principali definizioni assunte alla base del progetto.

**Famiglia:** insieme di persone legate da vincoli di matrimonio, parentela, affinità, adozione, tutela o affettivi, coabitanti ed aventi dimora abituale nello stesso comune (anche se non sono ancora iscritte nell'anagrafe della popolazione residente del comune medesimo). Una famiglia può essere costituita anche da una sola persona. L'assente temporaneo non cessa di appartenere alla propria famiglia sia che si trovi presso altro alloggio (o convivenza) dello stesso comune, sia che si trovi in un altro comune. La definizione di famiglia (adottata anche per il Censimento della Popolazione) è quella contenuta nel Regolamento Anagrafico (legge 24 Dicembre 1954, n. 1228; DPR 30 Maggio 1989, n.223).

**Famiglia fiscale:** nucleo familiare identificabile attraverso i dati fiscali. A differenza della famiglia anagrafica e dei suoi eventuali nuclei, costituiti dai soggetti conviventi nella stessa abitazione, la famiglia fiscale è definita in base alle norme del Tuir e individuata sulla base delle informazioni contenute nelle dichiarazioni dei redditi a fini fiscali. Essa risulta pertanto costituita dal contribuente dichiarante, dall'eventuale coniuge, dichiarante o meno, e da tutti i familiari fiscalmente a carico, indipendentemente dalla effettiva convivenza nella medesima dimora.

**Reddito:** definito come l'entrata (in termini monetari e non) realizzata da un soggetto in un determinato periodo di tempo. Rappresenta il divenire di componenti economici attribuito ad un dato periodo di tempo. Il reddito è quindi una variabile di flusso, mentre il concetto di patrimonio esprime in termini monetari la ricchezza in un dato istante. È costituito dalla somma di redditi provenienti da fonti diverse:

- lavoro dipendente;
- pensione;
- lavoro autonomo (reddito agrario, allevamento, reddito da lavoro autonomo, altri redditi di lavoro autonomo, provvigioni e redditi diversi, altri redditi diversi);
- impresa (reddito d'impresa in contabilità ordinaria, reddito d'impresa in contabilità semplificata);
- partecipazione;
- fabbricati;
- capitale (reddito dominicale, redditi da capitale, plusvalenze, altri redditi da capitale).

**Reddito equivalente:** Per confrontare i redditi delle famiglie è opportuno tener conto del fatto che le famiglie più numerose hanno maggiori bisogni. Il reddito pro-capite ignora completamente le “economie di scala”, cioè il fatto che i costi di una famiglia non sono perfettamente proporzionali al numero di componenti. In ambito internazionale, questo limite è stato superato utilizzando una scala di equivalenza, cioè un insieme di parametri che tengono conto della diversa composizione delle famiglie. Per calcolare il reddito “equivalente” possono essere utilizzate delle scale di equivalenza; la più famosa è la scala OCSE modificata da EUROSTAT, che assegna un peso pari a 1 al primo adulto, 0,5 agli altri adulti che compongono il nucleo familiare, e 0,3 per ogni membro al di sotto dei 14 anni. La costruzione del coefficiente per cui viene diviso il reddito familiare è quindi vincolato alla disponibilità di informazioni puntali associabili ai singoli componenti del nucleo familiare. Le informazioni contenute nella base dati integrata consentono di calcolare il reddito familiare equivalente, il quale sarà utilizzato per effettuare confronti tra il reddito delle famiglie, indipendentemente dalla loro struttura.

**Intensità del lavoro:** è convenzionalmente definita come il rapporto fra il numero totale di mesi lavorati dai componenti della famiglia durante l’anno di riferimento dei redditi (quello precedente all’anno di rilevazione) e il numero totale di mesi teoricamente disponibili per attività lavorative. Ai fini del calcolo di tale rapporto, come indicato da Eurostat, si considerano i membri della famiglia di età compresa fra i 18 e i 59 anni, escludendo gli studenti nella fascia di età tra i 18 e i 24 anni. Famiglie composte soltanto da bambini, da studenti di età inferiore a 25 anni e/o persone di 60 anni o più non vengono considerate nel calcolo dell’indicatore.

#### 4.1.3 Le fonti dei dati

La ricognizione delle informazioni necessarie a coprire le dimensioni anzidette e a calcolare i relativi indicatori, ha permesso di individuare le fonti di natura amministrativa e censuaria, e le variabili che costituiscono “*tracking tools*” di base per sviluppare il progetto.

Ad eccezione del Censimento della Popolazione, si tratta di fonti di dati amministrativi acquisite dall’Istat per vari processi produttivi, che contengono informazioni riferite a popolazioni diverse (es. studenti, tipologie specifiche di lavoratori, pensionati, ecc.). In generale ciascuna fonte amministrativa acquisita è pre-trattata e sottoposta ad analisi e validazione.

Alcune fonti, individuate come potenzialmente utili al progetto, sono ancora in fase di studio e la loro acquisizione verrà inserita per le fasi di sviluppo successivo del progetto.

Nel prospetto 4.1 sono elencate le fonti valutate di interesse per il progetto, sia quelle disponibili sia quelle non ancora acquisite e sono specificati i contenuti informativi di riferimento per il progetto. Si ritiene utile approfondire la disponibilità dell’Anagrafe dei conti correnti e dei dati del Catasto. La prima è un’anagrafe in fieri, cui l’Agenzia delle Entrate sta dando un forte impulso; una volta entrata a regime le informazioni deducibili da questa fonte potranno completare il quadro della ricchezza a disposizione degli individui e delle famiglie. Nella stessa direzione la seconda fonte, il Catasto,<sup>48</sup> potrà fornire preziose informazioni relative a quei proprietari di case che non sono individuabili tramite le dichiarazioni dei redditi.

<sup>48</sup> Agenzie delle Entrate (2013).

### Prospetto 4.1 - Le fonti amministrative per il progetto precarietà delle condizioni socio-economiche delle famiglie

FONTI DI DATI	Disponibilità	Contenuti informativi utilizzabili nel progetto
Censimento Popolazione 2011	Disponibile in base ai tempi della validazione dei dati	Caratteristiche anagrafiche dei componenti
		Livello di Istruzione
		Informazioni su abitazioni
Lista Anagrafiche Comunali (LAC)	Disponibile	Caratteristiche anagrafiche dei componenti Dimensione familiare e tipo di famiglia
Quadro familiari a carico delle dichiarazioni reddituali	Disponibile	Familiari a carico (individuare se convivente o meno)
Banca dati Redditi (Unico/730/770)	Disponibile	Redditi delle persone/famiglie
Casellario dei pensionati	Disponibile	Pensionati
		Percettori di pensioni assistenziali → condizione di disabilità
Gestione fiscale per prestazioni non pensionistiche	Disponibile	Indennità di Disoccupazione/Mobilità/LSU LPU
Anagrafe degli studenti delle scuole	Disponibile	Livello di Istruzione
Anagrafe studenti universitari	Disponibile	Livello di Istruzione
Anagrafe AFAM	<i>Non disponibile</i>	Livello di Istruzione
Anagrafi regionali degli studenti	<i>Non disponibile</i>	Livello di Istruzione
Asia occupazione	Disponibile	Tipologia occupati, Numero, Tipologia e durata del contratto
INPDAP	Disponibile	Numero, Tipologia e durata del contratto
Archivio lavoratori domestici	Disponibile	Numero, Tipologia e durata del contratto
Autonomi in agricoltura	Disponibile	Tipologia di occupati
Archivio possessori social card	<i>Non disponibile</i>	Reddito - situazione economica del nucleo familiare
Permessi di soggiorno	Disponibile	Stranieri extracomunitari con permesso di soggiorno in scadenza
Archivio ISEE	<i>Non disponibile</i>	Reddito - situazione economica del nucleo familiare
Catasto	<i>Non disponibile</i>	Reddito - situazione economica del nucleo familiare; condizione abitativa
Anagrafe dei conti correnti	<i>Non disponibile</i>	Reddito - situazione economica del nucleo familiare

#### 4.1.4 Le attività sperimentali

Accertata l'effettiva disponibilità delle fonti e il loro allineamento temporale, la sperimentazione ha potuto contare su nove fonti di dati riferite al 2011. Selezionando le variabili contenute in tali fonti è stato possibile valorizzare 5 dimensioni su 6, ovvero tutte quelle definite ad eccezione della situazione abitativa.

Per ciascuna dimensione è stato definito un appropriato set di variabili utilizzando una o più fonti amministrative integrate. In particolare, si segnala che:

- la tipologia familiare è stata ricostruita utilizzando le variabili disponibili nelle Liste Anagrafiche Comunali (LAC) (cfr. § 4.2.2);
- il reddito fiscale è stato integrato con fonti informative relative a popolazioni complementari di percipienti di redditi esenti (cfr. § 4.2.3);
- il mercato del lavoro è stato esaminato limitatamente alla presenza occupazionale nel mese di ottobre, definendo sotto-popolazioni di lavoratori (cfr. § 4.2.4.3.);
- il disagio è stato esaminato utilizzando le informazioni del Casellario dei Pensionati sui percettori di pensioni di disabilità e di pensioni sociali, ma anche valorizzando l'informazione sulla nazionalità dei componenti il nucleo familiare (cfr. § 4.2.4.1);
- l'istruzione e la formazione sono state trattate con informazioni di tipo *flag* limitatamente a sotto-popolazioni degli studenti, ossia di chi risulta iscritto ad un percorso di studio (cfr. § 4.2.4.2).

La sperimentazione è stata effettuata per i comuni di Modena, Bologna e Firenze.

Il primo risultato conseguito della sperimentazione è stato quello di progettare e realizzare l'integrazione concettuale (metadati) e dei dati provenienti dalle diverse fonti, al fine di organizzare una base dati fruibile potenzialmente da utenti interni ed esterni all'Istat.

#### 4.2 I risultati

In questa sezione del documento si espongono i risultati della sperimentazione svolta nel corso del 2013. In particolare si presentano i contenuti informativi e la struttura degli archivi integrati di microdati prodotti e le principali procedure seguite per la loro realizzazione.

Valorizzando i bacini informativi presenti in Istat sono state prodotte collezioni di microdati sugli individui e sulle famiglie, con l'inserimento di variabili ritenute rilevanti, ottenute a partire da integrazione, elaborazione o sintesi di singole variabili disponibili nei singoli archivi.

Queste collezioni permettono di classificare le famiglie di un dato territorio in base alle loro condizioni socio-economiche e, in particolare, sono predisposte in maniera tale da consentire l'individuazione di differenti situazioni di criticità. A partire dai microdati è possibile inoltre calcolare vari indicatori, al fine di delineare dei profili socio-economici dei territori.

#### 4.2.1 I dati elementari: struttura e contenuti informativi

Per la sperimentazione sono stati progettati e realizzati dei set di dati elementari riferiti agli individui e alle famiglie residenti (all'08/10/2011) nei comuni di Modena, Bologna e Firenze. In particolare, per ognuno di questi comuni, sono state costruite due basi di dati aventi rispettivamente come unità il singolo individuo e la famiglia anagrafica.

Le principali informazioni contenute delineano le caratteristiche socio-economiche delle famiglie e sono:

- tipologia familiare; numero di componenti per sesso e classe di età;
- reddito familiare complessivo per macro-voci e apporto di ogni componente al reddito familiare;
- numero di componenti della famiglia con disabilità, per classe di età;
- classificazione della famiglia in base alla cittadinanza dei componenti;
- partecipazione di componenti della famiglia a percorsi scolastici ed universitari;
- partecipazione di componenti della famiglia al mercato del lavoro e caratteristiche dell'occupazione;
- presenza in famiglia di pensionati con reddito minimo.

#### 4.2.2 La popolazione di base e la costruzione della tipologia familiare

La popolazione di riferimento per la sperimentazione è quella degli iscritti nelle Anagrafi dei Comuni riferite all'8 ottobre 2011,<sup>49</sup> Si tratta complessivamente di 933.169 individui afferenti a 474.213 famiglie: Bologna è il comune più popoloso (378.847 individui residenti), seguito da Firenze (370.142 individui residenti) e Modena (184.180 individui residenti) (Tabella 4.1).

**Tavola 4.1 - Numero di individui e di famiglie residenti nei comuni della sperimentazione allo 08/10/2011**

COMUNE	N. individui	N. famiglie
Bologna	378.847	204.304
Firenze	370.142	185.265
Modena	184.180	84.644

Fonte: Liste Anagrafiche Comunali (LAC)

Le famiglie anagrafiche possono essere classificate in maniera diversa in base alle caratteristiche dei loro componenti (come lo stato civile, l'età, il sesso) e alle relazioni tra loro intercorrenti. Tuttavia nell'ambito della sperimentazione si è ritenuto opportuno adottare una classificazione standard, mutuata da quella del Censimento della popolazione e con un minor numero di modalità (5 invece di 17). Questa articolazione risulta essere l'unica possibile in virtù delle informazioni contenute nelle LAC. Si fa presente, tuttavia, che essa può essere integrata con i dati relativi al numero di componenti per età e per sesso; tali informazioni sono state infatti inserite nel dataset familiare al fine di distinguere meglio situazioni familiari diverse. Le modalità della variabile "tipologia familiare" che è stato possibile costruire attraverso la procedura implementata sono le seguenti:

<sup>49</sup> Si fa presente che nei comuni di Modena e Firenze ci sono dei record con codice fiscale non valorizzato. Sebbene essi siano inclusi nella popolazione di riferimento e siano conteggiati nelle elaborazioni presentate in questo paper, è in fase di valutazione l'ipotesi di considerare non eleggibili i record con codice fiscale mancante.

“unipersonale”; “coppie (coniugate e non) senza figli”; “coppie (coniugate e non) con figli”; “monogenitore”; “altro”. Le variabili delle LAC impiegate nel corso della procedura di classificazione sono la relazione di parentela (espressa in relazione all’intestatario del foglio di famiglia), lo stato civile e il numero di componenti. Ognuna delle cinque modalità della tipologia familiare è determinata dalla combinazione di alcune modalità della relazione di parentela e dello stato civile, sotto alcune condizioni. I risultati derivanti dell’algoritmo implementato sono stati messi a confronto con la distribuzione delle tipologie familiari pubblicate dai Comuni<sup>50</sup> (Tavola 4.2). In generale la distribuzione in tutti e tre i comuni mostra un andamento corrispondente, tuttavia si rileva una maggiore incidenza della modalità residuale “altro” come esito della procedura eseguita per la sperimentazione a scapito della modalità “monogenitore” nel comune di Bologna, della modalità “coppie senza figli” nel comune di Modena e delle modalità “coppie con figli” “monogenitore” nel comune di Firenze (Tabella 4.3).

**Tavola 4.2 - Confronto esito procedura classificazione con dati di benchmark prodotti dagli uffici comunali**

BOLOGNA	Da procedura		Da dati benchmark (a)	
	N Famiglie	Percentuale	N Famiglie	Percentuale
Unipersonale	103.117	50,47	103.513	50,59
Coppie senza figli	31.863	15,60	31.791	15,54
Coppie con figli	31.354	15,35	32.228	15,75
Monogenitore	18.267	8,94	20.225	9,89
Altro	19.703	9,64	16.841	8,23
<b>Totale</b>	<b>204.304</b>	<b>100</b>	<b>204.598</b>	<b>100</b>

MODENA	Da procedura		Da dati benchmark (b)	
	N Famiglie	Percentuale	N Famiglie	Percentuale
Unipersonale	32.881	38,85	32.903	38,84
Coppie senza figli	14.455	17,08	17.127	20,22
Coppie con figli	17.838	21,07	18.497	21,83
Monogenitore	6.937	8,20	7.462	8,81
Altro	12.533	14,81	8.725	10,3
<b>Totale</b>	<b>84.644</b>	<b>100</b>	<b>84.714</b>	<b>100</b>

FIRENZE	Da procedura		Da dati benchmark (c)	
	N Famiglie	Percentuale	N Famiglie	Percentuale
Unipersonale	85.656	46,23	84.517	45,92
Coppie senza figli	26.857	14,50	27.020	14,68
Coppie con figli	32.394	17,49	37.256	20,24
Monogenitore	18.161	9,80	26.250	14,26
Altro	22.197	11,98	9.000	4,89
<b>Totale</b>	<b>185.265</b>	<b>100</b>	<b>184.043</b>	<b>100</b>

(a) Fonte: [http://www.comune.bologna.it/iperbole/piancont/dati\\_statistici/Indici/Popolazione/index.htm](http://www.comune.bologna.it/iperbole/piancont/dati_statistici/Indici/Popolazione/index.htm). Dati al 2011. Nota: le tipologie familiari non sono del tutto corrispondenti con quelle delle Lac, in particolare le coppie non coniugate ricadono in altre tipologie familiari.

(b) Fonte: <http://sis.comune.modena.it/SaSWeb/faces/index.jsp>.

(c) Dati comunali 28 febbraio 2011. Fonte: [http://statistica.fi.it/opencms/multimedia/documents/1300796401765\\_bollettino\\_marzo31.pdf](http://statistica.fi.it/opencms/multimedia/documents/1300796401765_bollettino_marzo31.pdf).

<sup>50</sup> La data di riferimento di tali dati non coincide con quella di LAC; tuttavia entrambe fanno riferimento al 2011.

### 4.2.3 Il reddito familiare

Uno degli obiettivi della sperimentazione consiste nella ricostruzione del reddito complessivo lordo delle famiglie inteso come somma di una serie di componenti di fonte diversa, indipendentemente dal fatto che siano tassate o esenti e dall'eventuale tipologia di tassazione (ordinaria, separata).

In sintesi gli obiettivi conoscitivi ottimali del progetto sono:

1. calcolo del reddito complessivo (non solo fiscale) della famiglia;
2. stima del livello di partecipazione dei componenti della famiglia alla formazione del reddito (anche in relazione alla loro fascia di età);
3. computo delle singole voci del reddito complessivo (es. da lavoro, da pensione, da capitale, ecc.).

L'attività di ricognizione delle fonti conosciute e disponibili in Istat ha permesso di individuare alcune componenti del reddito calcolabili e di approntare una classificazione in macro-categorie, che ricalca quella adottata per Eu-Silc;<sup>51</sup> questa rappresenta, a sua volta, un adattamento al contesto italiano di quella internazionale riportata nel manuale di Canberra.<sup>52</sup>

La ricostruzione e la classificazione del reddito implica un processo metodologico con un elevato livello di complessità che, inoltre, deve consentire l'uso di altri sistemi di classificazione adottati nei processi produttivi dell'Istituto.

Per la sperimentazione realizzata in questi mesi sono state valorizzate le informazioni contenute negli archivi disponibili in Istat e allineati al 2011, che sono:

- la Banca Dati Reddittuale (BDR): si tratta di una base statistica realizzata dal Ministero dell'Economia e delle Finanze (MEF) a partire dalle informazioni contenute nelle dichiarazioni dei redditi effettuate dai contribuenti (o dai sostituti d'imposta) attraverso i modelli Unico, 730 e 770;
- il Casellario dei Pensionati (CP): contiene i trattamenti pensionistici obbligatori ed integrativi erogati da vari enti e comunicati periodicamente all'Inps;
- l'archivio dei lavoratori domestici dell'Inps: include i microdati sui rapporti di lavoro domestico attivi nell'anno di riferimento e sui relativi contributi trimestrali versati dal datore di lavoro;
- il quadro Contribuenti Minimi (CM) dell'Agenzia delle Entrate (quadro presente nella dichiarazione Unico PF ma assente in BDR): contiene l'informazione puntuale sul reddito delle persone fisiche esercenti attività di impresa, arti o professioni che hanno aderito al regime contabile agevolato poiché in possesso dei seguenti requisiti: hanno un tetto massimo di ricavi nell'anno precedente (non superiori a 30.000 euro) e nel triennio solare precedente non hanno effettuato acquisti di beni strumentali (per un ammontare complessivo superiore a 15.000 euro) e al contempo non hanno effettuato esportazioni, né hanno erogato somme sotto forma di utili di partecipazioni agli associati con apporto costituito da solo lavoro, né hanno sostenuto spese per lavoratori dipendenti o collaboratori.

Tali procedure potranno essere perfezionate a seguito delle evidenze che emergeranno dalle attività del gruppo di lavoro interdipartimentale solo di recente costituito in Istat per l'analisi della BDR.<sup>53</sup>

Alcuni punti di partenza devono essere evidenziati:

- il reddito ricostruito in questo progetto non è derivato in nessuna sua componente da indagini ma solo attraverso l'integrazione di archivi amministrativi di natura fiscale e contributiva;
- il reddito è riferito a tutta la popolazione residente nei comuni in esame all'8/10/2011;
- la sotto-popolazione cui è stata associata una qualsiasi quota di reddito è definita "percettori di reddito" ed è numericamente più consistente della popolazione dei "dichiaranti";
- non sono incluse nel computo del reddito alcune poste, sia tra quelle assoggettate a imposizione

<sup>51</sup> Istat (2011a).

<sup>52</sup> United Nations (2011).

<sup>53</sup> Cfr. Delibera DGEN n. 173 del 4 Novembre 2013.



sostitutiva sia tra quelle per cui non sussistono obblighi dichiarativi né implicazioni di natura contributiva.

La dimensione del reddito è cruciale per definire e studiare le condizioni socio-economiche di un individuo e della sua famiglia. Lo sfruttamento dei dati amministrativi consente di ricostruire il “flusso di reddito” generato nell’anno, rappresentato principalmente dalla retribuzione percepita per le attività di lavoro (sia autonome sia dipendente) o da pensione.

Al contrario le componenti di reddito derivanti da cespiti patrimoniali non possono essere ricostruite attraverso i dati attualmente a disposizione, se non in parte limitata e come proxy. È quest’ultimo il caso dei redditi da capitale reale, in particolare quelli da fabbricati, che possono essere individuati solo parzialmente per la sotto-popolazione dei contribuenti che compilano il modello Unico o il 730.

Le variabili inserite nella base dati individuale del sotto-progetto consentono di stimare le seguenti componenti del reddito:

1. **redditi da lavoro dipendente e assimilati.** Tra le voci in cui può essere scomposto il reddito da lavoro dipendente, il dato relativo agli stipendi ne costituisce la quota principale. Gli archivi amministrativi contengono questa voce di reddito (comprese le borse di studio non esenti erogate al di fuori di un rapporto di lavoro) così come alcune integrazioni monetarie quali i premi e lo straordinario. Tuttavia non si dispone delle altre integrazioni monetarie (ad esempio provvigioni e gratifiche, partecipazioni e titoli, ecc.) e non monetarie, quali i fringe benefits. La popolazione di riferimento è eterogenea e comprende oltre ai lavoratori dipendenti in senso stretto, altre tipologie di percettori quali: i lavoratori soci di cooperative di produzione e lavoro, i lavoratori socialmente utili, i co.co.co, i co.pro, sacerdoti e altri ministri di culto;
2. **redditi da lavoro autonomo.** I redditi da lavoro autonomo sono stati scomposti in più voci associate a diverse tipologie di percettori: titolari di un’attività agricola o di un’impresa agricola individuale, chi svolge un’attività di allevamento di animali e/o dirette alla produzione di vegetali, lavoratori autonomi che esercitano arti e professioni, coloro che svolgono attività sportive dilettantistiche o collaborazioni con cori bande o filodrammatiche, autori o inventori che realizzano proventi relativi alla cessione di opere e invenzioni, tutelate dalle norme sul diritto d’autore anche in via occasionale (brevetti, disegni e modelli ornamentali e di utilità, know-how, articoli per riviste o giornali, ecc.), segretari comunali che esercitano attività di levata dei protesti, associati in partecipazione, titolari di impresa, coadiuvanti in impresa familiare e soci di società di persone. Tra le voci di reddito autonomo si possono annoverare anche le indennità percepite per la cessazione di rapporti di agenzia, da funzioni notarili o percepite da sportivi professionisti al termine dell’attività sportiva che sono compresi tra i redditi a tassazione separata attualmente non disponibili;
3. **redditi da capitale reale.** Le voci classificate nella categoria “capitale reale”, individuate nelle fonti disponibili, consentono di calcolare solo parzialmente i redditi derivanti dal possesso di un immobile adibito ad abitazione principale (trattasi di reddito figurativo) o in locazione e i redditi dominicali. Sarebbe opportuno completare le informazioni acquisendo ulteriori archivi esistenti a livello nazionale o locale, o comunque valutando, sia in termini di popolazione (possessori di terreni e fabbricati esonerati dalla dichiarazione dei redditi) sia in termini economici, quanto impatta l’informazione mancante sul reddito totale. Di contro è possibile calcolare i redditi da fabbricati in locazione in regime di tassazione separata (cedolare secca);
4. **redditi da capitale finanziario.** Nella categoria “capitale finanziario” sono compresi i redditi derivanti dalla partecipazione al capitale di società ed enti soggetti all’Ires e quelli distribuiti da società ed enti esteri di ogni tipo, unitamente a tutti gli altri redditi di capitale, percepiti nel 2011, senza avere riguardo al momento in cui è sorto il diritto a percepirli e i redditi trasferiti al dichiarante da Trust trasparenti o misti di cui all’art. 73, comma 2, del TUIR di cui lo stesso è beneficiario;
5. **redditi da pensione.** L’integrazione dei dati BDR e CP consente di fornire informazioni sull’importo di diverse pensioni, sia di quelle a tassazione ordinaria, sia di quelle esenti;

6. **redditi da altri trasferimenti pubblici.** In questa voce possono incluse una serie di prestazioni, di natura sociale e assistenziale, a sostegno delle famiglie o a particolari categorie di soggetti svantaggiati. Tuttavia in questa fase sperimentale, data la mancata disponibilità di alcune fonti amministrative Inps, tale voce comprende solo il reddito pensionistico derivante dalla/e pensione/i di invalidità civile. Le pensioni sociali e di invalidità civile sono state classificate come redditi da trasferimenti pubblici e non come redditi da pensione in quanto esse si configurano come prestazioni di carattere assistenziale che prescindono del tutto dal versamento dei contributi e spettano ai cittadini che si trovino in condizioni economiche disagiate o che per minorazioni congenite o acquisite, siano incapaci di svolgere un'attività lavorativa;
7. **redditi da altri trasferimenti privati.** È disponibile esclusivamente l'assegno ricevuto da coniuge separato o divorziato;

Le singole voci di reddito dei componenti la famiglia confluiscono, come somma, nella variabile reddito complessivo. Questo ricomprende elementi reddituali presenti in banche dati amministrative differenti. Inoltre, l'articolazione delle voci di reddito consente di misurare il peso di ciascuna componente sul totale, sia per unità individuo che per unità famiglia.

Il **valore informativo che deriva dall'integrazione di fonti amministrative** è di assoluto rilievo in quanto le informazioni derivabili dalla Banca Dati Reddittuale del MEF non sono sufficienti a rilevare tutte le possibili voci di reddito. Sebbene l'integrazione non permetta di coprire ogni singola componente del reddito, essa consente di includere alcune voci che, se non considerate, potrebbero implicare una sottostima dell'ammontare e una distorsione nella classificazione del reddito familiare. A titolo esemplificativo, si fa presente che, nel comune di Modena, l'integrazione di fonti diverse rispetto alla sola BDR, permette di incrementare il numero di percettori di reddito del 4,05% e l'ammontare del reddito del 3,19% (soprattutto per le macro voci reddito da trasferimenti pubblici e reddito da lavoro autonomo).

#### *4.2.4 Le sottopopolazioni individuabili e i loro attributi*

Le collezioni di microdati sono state predisposte in maniera tale da consentire l'individuazione di alcune particolari sottopopolazioni di soggetti relativamente alle dimensioni del disagio, dell'istruzione e del mercato del lavoro. In questo paragrafo si presentano le principali sottopopolazioni, si dà conto delle modalità attraverso le quali esse sono state caratterizzate e si espongono i principali attributi ad esse riferibili.

##### *4.2.4.1. Le sottopopolazioni afferenti alla dimensione Disagio*

Nell'ambito della dimensione del disagio vengono ricomprese varie categorie di persone. I fattori di disagio investono diversi ambiti della vita degli individui e delle famiglie e possono essere associati ad una probabilità di povertà futura: la discriminazione, l'emarginazione, l'esclusione sulla base della cittadinanza, del credo religioso, del genere, della disabilità fisica, sono tutti elementi che possono avere impatto sul benessere di una famiglia.

La caratterizzazione delle persone disagiate adottata in questo studio è ovviamente vincolata alla disponibilità delle informazioni contenute all'interno degli archivi amministrativi. La fase preliminare di studio della struttura e del contenuto informativo delle variabili originarie ha permesso di individuare delle sotto-popolazioni di interesse che sono: le persone che percepiscono pensioni di disabilità, i percettori di indennità di accompagnamento, le persone con una pensione sociale, le famiglie straniere.

##### *I percettori di pensioni di disabilità*

I dati disponibili hanno permesso di distinguere i percettori di pensioni a cui è possibile associare una condizione di disabilità del titolare. Le informazioni di input sono derivate dal Casellario dei pensionati sulle prestazioni pensionistiche riconosciute a:

1. invalidi civili (individui che abbiano subito una riduzione permanente della capacità lavorativa non inferiore ad un terzo o, se minori di anni 18, che abbiano difficoltà persistenti a svolgere i compiti e le funzioni proprie della loro età, art. 2 l.118/1971);
2. invalidi di guerra;
3. invalidi del lavoro e invalidi per servizio (rispettivamente per i lavoratori privati o per i dipendenti pubblici e le categorie assimilate);
4. ciechi e sordomuti.

La procedura per l'individuazione dei percettori di prestazioni di disabilità<sup>54</sup> consta di due parti: la prima, effettuata sul dataset dei percettori di prestazioni pensionistiche, è finalizzata all'identificazione delle varie tipologie di prestazioni di invalidità; la seconda è volta a riportare tale informazione dall'unità 'prestazione' all'unità 'percettore'.

In particolare, l'attuazione della prima parte della procedura porta, attraverso la combinazione di modalità di variabili del Casellario dei pensionati, alla creazione di una nuova variabile, che permette di identificare le seguenti tipologie di prestazioni riconducibili ad una situazione di disabilità:

- a) pensioni di invalidità: sono erogate a coloro che hanno lavorato almeno 3 anni negli ultimi 5 anni di riferimento. Sono identificate dal verificarsi della combinazione: titolarità diretta e tipologia di pensione invalidità;
- b) pensioni indennitarie: pensioni erogate in seguito a malattie professionali e infortuni sul lavoro. Sono individuate da: titolarità diretta e tipologia di pensione indennitaria;
- c) pensioni di invalidità civile, individuate dalle seguenti modalità della variabile tipologia di prestazione assistenziale: pensioni di invalidità civile, pensioni ai non udenti civili e pensioni ai non vedenti civili. Confluiscono inoltre in questa modalità le pensioni sociali che prima dei 65 anni erano pensioni di invalidità civile. Quest'ultima tipologia di pensione è individuata dalla combinazione: tipologia di prestazione assistenziale = 'pensione o assegno sociale' e gestione inps = 'pensione sociale e assegno sociale da invalidità civile';
- d) pensioni di guerra: sono erogate a coloro che hanno subito menomazioni derivanti da eventi bellici.

La seconda parte della procedura riconduce ai percettori questa informazione riferita alla prestazione. Per ogni individuo associato ad una o a più prestazioni viene creata una variabile che presenta le seguenti modalità:

1. = invalidità;
2. = indennitarie;
3. = invalidità civile;
4. = guerra
5. = cumulo (più prestazioni di disabilità).

Per ogni individuo è possibile, dunque, identificare il tipo di prestazione di invalidità percepita, nel caso in cui sia beneficiario di una sola prestazione legata alla disabilità oppure se percepisce più di una pensione associata a disabilità.

Questa informazione tuttavia viene riportata nel dataset individuale finale soltanto come flag, il quale consente di distinguere gli individui che non risultano percettori di pensioni di disabilità dagli individui che risultano percettori di una o più pensioni di disabilità.

Successivamente le informazioni vengono trasposte dal livello individuale al livello familiare. Per ogni famiglia si dispone dunque del numero di percettori di pensioni di disabilità e del numero di percettori di pensioni di disabilità per classe di età.

#### *I percettori di indennità di accompagnamento*

Le informazioni contenute nel Casellario dei Pensionati consentono inoltre di individuare, tra i percettori di prestazioni di disabilità, coloro i quali si vedono riconosciuta una invalidità totale e

<sup>54</sup> La procedura è stata curata dalla DCSE/CEF/I U.O. Statistiche sul sistema pensionistico.

permanente del 100%. Si tratta di individui che non possono deambulare senza l'aiuto permanente di un accompagnatore o che, non essendo in grado di compiere gli atti quotidiani della vita, necessitano di un'assistenza continua; per questo motivo i titolari di tali indennità accessorie possono essere considerati come soggetti affetti da invalidità maggiore. I percettori di invalidità di accompagnamento vengono individuati grazie alla variabile del Casellario Pensionati che contiene l'importo mensile dell'indennità di accompagnamento; se un soggetto è titolare di almeno una prestazione per la quale risulta valorizzata questa variabile, egli viene classificato come percettore di indennità di accompagnamento.

#### *I percettori di pensione sociale*

Sono stati individuati i titolari di pensioni o assegni sociali e, tra di essi, coloro che ricevono unicamente pensioni di questo tipo, ritenendoli soggetti ad un rischio di povertà maggiore rispetto a coloro che cumulano questo trattamento ad altri. Questa ipotesi è avvalorata dai dati riferiti al 2010<sup>55</sup> dai quali si deduce che l'importo medio annuo dei redditi pensionistici dei beneficiari di due o più prestazioni è pari a più del doppio di quelli attribuiti ai pensionati che ricevono solo la pensione o l'assegno sociale.

#### *Le famiglie straniere*

Come emerge da numerosi studi, le famiglie straniere mostrano maggiori elementi di disagio rispetto alle famiglie italiane e, in particolare, dispongono di un reddito nettamente inferiore alle famiglie miste o italiane.<sup>56</sup>

Si è quindi ritenuto opportuno inserire un'informazione utile a classificare le famiglie residenti in base alla cittadinanza dei loro componenti. La variabile di partenza utilizzata per questa classificazione è rappresentata dalla cittadinanza dei singoli individui, tratta dalle LAC. Sul dataset familiare viene creata una variabile che permette di distinguere le famiglie con tutti i componenti con cittadinanza italiana, le famiglie con tutti i componenti aventi cittadinanza straniera e le famiglie miste.

#### **4.2.4.2. Le sottopopolazioni afferenti alla dimensione Istruzione**

Lo studio della vulnerabilità sociale comprende anche l'approfondimento di aspetti quali il mancato accesso all'istruzione, dal momento che tale facoltà si configura come un fattore di protezione dal rischio di povertà ed esclusione sociale.

Nell'ambito del progetto sperimentale è stato possibile individuare le sottopopolazioni degli studenti della scuola primaria, secondaria di I grado, secondaria di II grado e degli studenti universitari. Nei dataset vengono quindi proposte alcune variabili che fanno riferimento alla partecipazione dei componenti delle famiglie a diversi percorsi di istruzione.

#### *Gli studenti della scuola primaria e secondaria*

Per l'individuazione delle sottopopolazioni degli studenti della scuola primaria e secondaria ci si è basati sulle informazioni desumibili dagli archivi del Miur: Anagrafe degli Studenti e Anagrafe delle Scuole.

Le informazioni prese in considerazione si riferiscono all'anno di iscrizione e al tipo di scuola frequentata (primaria, secondaria di I grado, secondaria di II grado). L'estrazione di queste informazioni è avvenuta vincolandosi ai codici fiscali degli individui residenti nei comuni della sperimentazione. Poiché nell'Anagrafe degli Studenti possono essere presenti diversi record associati ad uno stesso individuo, ad esempio coloro che cambiano scuola nel corso dell'anno, è stato necessario individuare dei criteri di deduplicazione. In particolare, nel caso di duplicazioni, sono stati sele-

<sup>55</sup> Istat (2012).

<sup>56</sup> Istat (2011b).

zionati i record che risultassero essere frequentanti (con conseguente esclusione dei soggetti che hanno abbandonato la scuola o che sono stati trasferiti). Si ritiene opportuno effettuare in futuro una revisione della procedura di individuazione della sottopopolazione di studenti anche nel caso in cui ad un individuo sia associato solamente un record.

#### *Gli studenti universitari*

Le informazioni sugli studenti universitari sono desunte dal Sistema Informativo *Persons&Places*, al quale si rimanda per le specifiche tecniche. Le variabili disponibili sul dataset individuale dopo la procedura di linkage permettono di individuare gli studenti universitari. Si è ritenuto opportuno, nel dataset familiare, distinguere tra gli studenti universitari di età inferiore ai 35 anni e quelli di età superiore ai 35 anni.

#### **4.2.4.3. Le sottopopolazioni afferenti alla dimensione Lavoro**

La partecipazione al Mercato del Lavoro è un tema cruciale negli studi sulle condizioni economiche e sociali delle famiglie. Lo stato occupazionale, infatti, è correlato positivamente con la disponibilità economica e quindi con la creazione di reddito e di sostentamento. Le caratteristiche professionali (tipo di lavoro, contratto, status) e occupazionali (durata del lavoro, ecc.) rappresentano elementi informativi indispensabili per effettuare analisi approfondite sullo stato di precarietà socio-economica di una famiglia. A tale scopo sono state associate agli individui residenti nei tre comuni oggetto di questa sperimentazione alcune delle variabili calcolate nell'ambito del sotto-progetto "Precarietà Lavorativa".

#### *Gli occupati nel mese di ottobre 2011*

La procedura di integrazione degli individui residenti nei tre comuni della sperimentazione con la popolazione degli occupati nel mese di ottobre 2011 costruita per il sotto-progetto "Precarietà Lavorativa" ha consentito di individuare tutti i soggetti che, nel periodo di riferimento, hanno svolto un'attività lavorativa per almeno un giorno.

#### *I precari nel mese di ottobre 2011*

La sotto-popolazione individuata è costituita dagli occupati che nel periodo di riferimento posseggono esclusivamente contratti di lavoro definiti precari, anche svolti contemporaneamente. I criteri per l'individuazione di questa sottopopolazione sono stati mutuati dal sistema "Precarietà Lavorativa".

### **4.3 Le potenzialità informative della base dei dati**

Elaborando le collezioni di dati, sia quella riferita agli individui che quella riferita alle famiglie, sono state prodotte tavole e grafici per descrivere e analizzare diverse le condizioni sociali ed economiche dei residenti nei comuni oggetto della sperimentazione.

Di seguito si presentano solo alcune delle possibili elaborazioni realizzate sperimentalmente a partire dai dati contenuti nei dataset. Ulteriori variabili ed indicatori possono essere realizzati "on demand" sulla base degli obiettivi conoscitivi dei possibili utilizzatori interni o esterni all'Istituto.

#### **4.3.1 Tavole dei dati**

Le elaborazioni sono organizzate in tre sezioni e mostrano mano a mano il valore aggiunto dell'informazione derivabile dall'integrazione delle fonti amministrative: nella prima sezione vi sono i dati sulle dimensioni reddito e famiglie; nella seconda sezione si aggiungono le informazioni che riguardano il disagio; nella terza sezione si evidenzia l'integrazione delle dimensioni istruzione e lavoro.

Per quanto riguarda la prima sezione, il grafico 4.1 presenta, tramite un box plot, la distribuzione del reddito familiare nei tre Comuni della sperimentazione, dalla quale si evince che il reddito

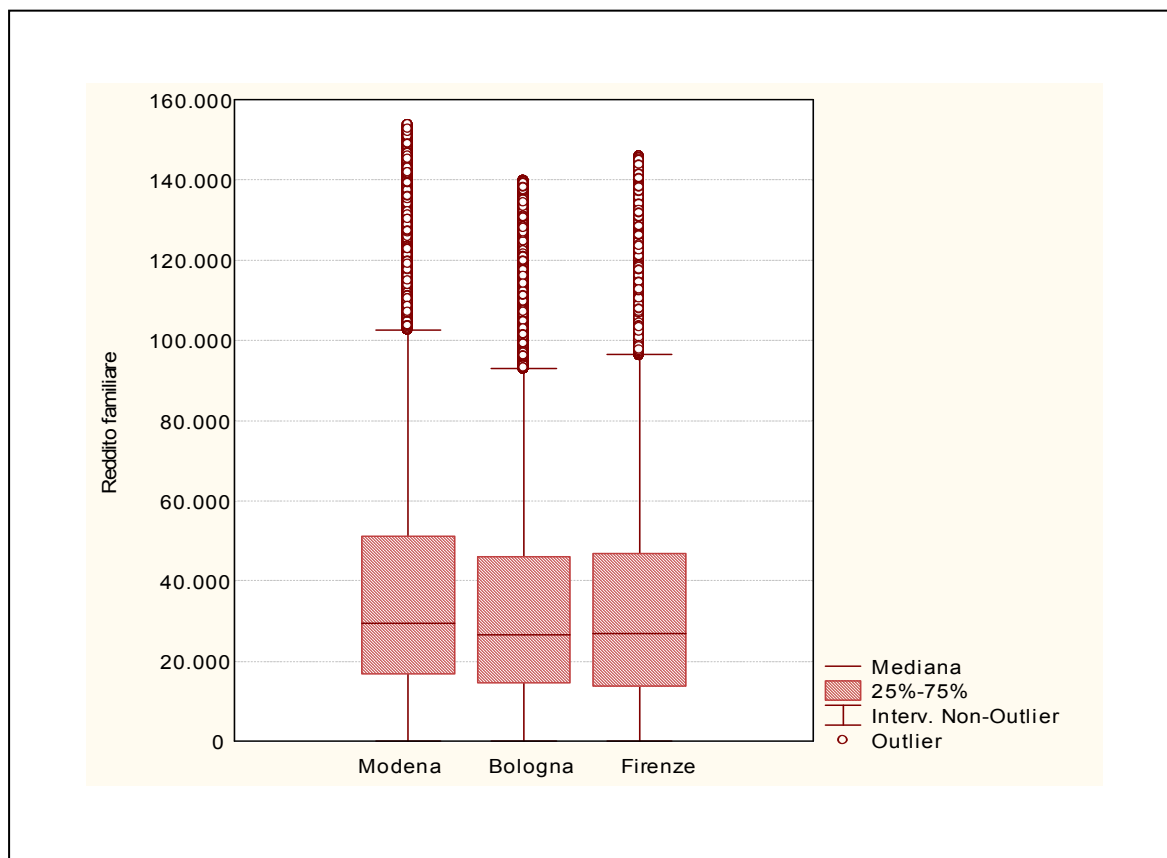
mediano più elevato è percepito nel Comune di Modena (si fa presente che questo risultato dipende tuttavia anche dal peso della composizione familiare). La tavola 4.3 riporta invece, per i tre comuni, i dati relativi al numero di famiglie, al reddito familiare medio, alla quota di reddito familiare da lavoro e alla quota di reddito familiare da pensione, per tipologia familiare e numero di percettori.

La seconda sezione include il grafico 4.2 e la tavola 4.4. Il grafico 4.2, mostra la composizione media del reddito per classe di età e condizione di disabilità. Si fa presente che la classe di età 0-14 anni è riportata al fine di evidenziare come siano presenti alcuni percettori di reddito con meno di 15 anni di età (si tratta in larga parte di percettori di pensioni di disabilità). Inoltre, si sottolinea che, sebbene siano presenti percettori di reddito da lavoro dipendente in età non lavorativa (inferiore ai 15 anni), essi sono numericamente esigui e ammissibili da un punto di vista normativo. La tavola 4.4 permette di confrontare, per tipologia familiare, il reddito mediano delle famiglie con membri con disabilità con il reddito mediano delle famiglie senza membri con disabilità.

La terza sezione include la tavola 4.5 che riporta il numero di famiglie con nessuna, una, due o più persone in età attiva (15-64) che non studiano, non lavorano senza reddito, per numero di componenti. Da essa si evince che la quota di persone impegnate in percorsi di istruzione, di lavoro o che percepiscono un reddito si attesta sopra l'84% in ogni comune e che essa varia sensibilmente in funzione del numero dei componenti della famiglia.

### #Reddito e famiglie

**Grafico 4.1 - Box plot del reddito familiare nei tre Comuni della sperimentazione (\*). Anno 2011. (Valori in euro)**



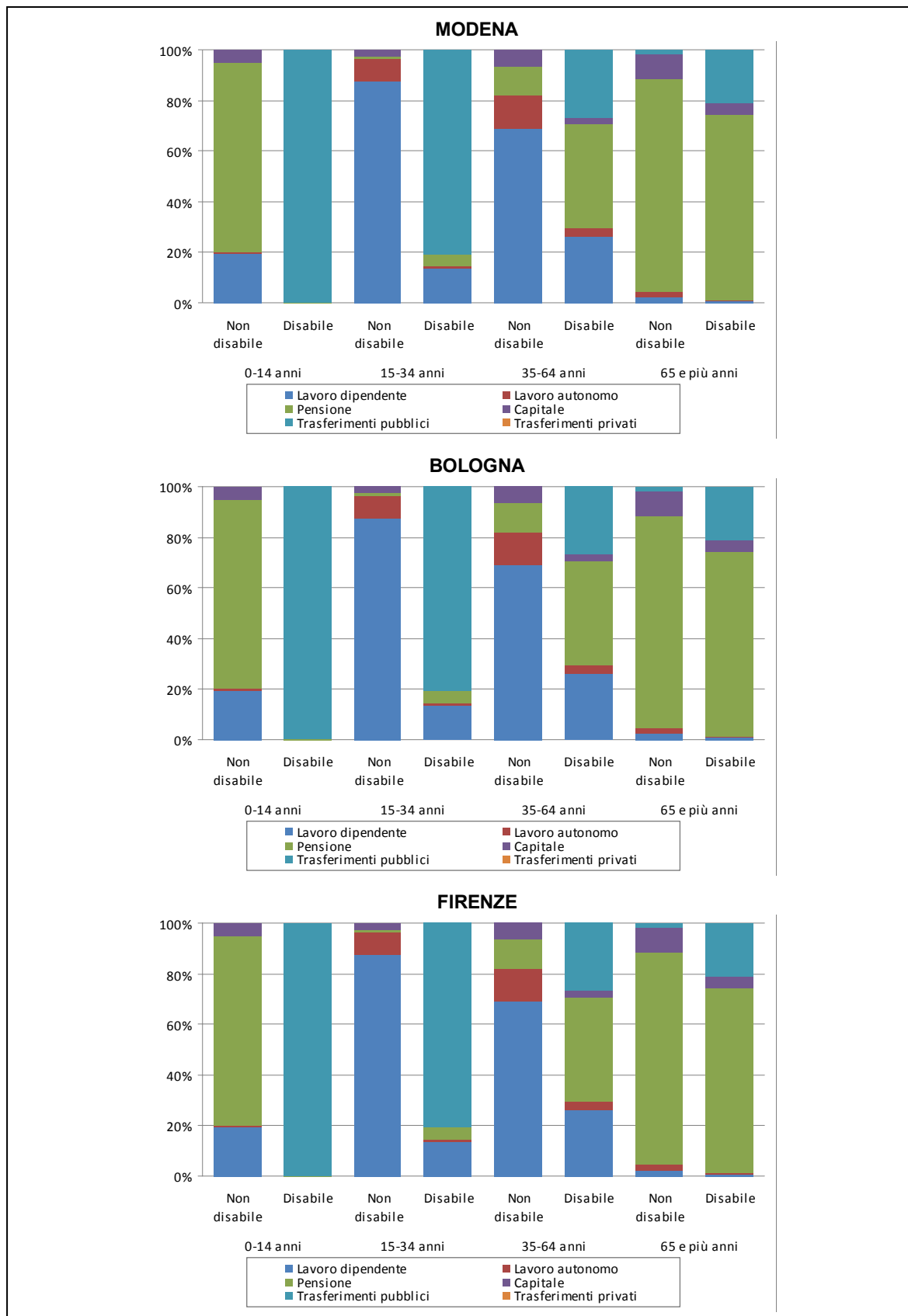
(\*) Per una maggiore visibilità del grafico sono stati esclusi gli outlier con valori di reddito superiori ai 160mila euro.

**Tavola 4.3 - Numero di famiglie, reddito familiare medio, quota di reddito familiare da lavoro e quota di reddito familiare da pensione, per tipologia familiare e numero di percettori. Anno 2011.**  
 (Reddito medio in euro)

<b>Modena</b>				
TIPOLOGIA FAMILIARE	N famiglie	Reddito familiare medio	% reddito da lavoro	% reddito da pensione
SOLO 1 PERCETTORE IN FAMIGLIA				
Unipersonale	30.042	24.754	55,99	34,88
Coppie senza figli	1.549	30.945	40,53	52,88
Coppie con figli	1.888	30.046	91,30	4,43
Monogenitore	3.082	27.558	85,22	7,75
Altro	2.015	19.050	86,26	9,73
DUE O PIÙ PERCETTORI IN FAMIGLIA				
Unipersonale	-	-	-	-
Coppie senza figli	12.816	50.789	36,18	53,91
Coppie con figli	15.764	69.202	82,96	11,49
Monogenitore	3.586	46.941	62,36	28,75
Altro	10.120	52.731	71,00	22,60
TOTALE				
Unipersonale	30.042	24.754	55,99	34,87
Coppie senza figli	14.365	48.649	36,47	53,84
Coppie con figli	17.652	65.014	83,37	11,15
Monogenitore	6.668	37.982	70,03	21,71
Altro	12.135	47.138	72,02	21,74
<b>Bologna</b>				
TIPOLOGIA FAMILIARE	N famiglie	Reddito familiare medio	% reddito da lavoro	% reddito da pensione
SOLO 1 PERCETTORE IN FAMIGLIA				
Unipersonale	91.105	25.917	58,14	32,35
Coppie senza figli	3.797	30.837	38,91	55,34
Coppie con figli	4.128	34.897	90,45	4,48
Monogenitore	8.966	29.783	83,10	7,92
Altro	2.925	20.232	71,30	20,73
DUE O PIÙ PERCETTORI IN FAMIGLIA				
Unipersonale	-	-	-	-
Coppie senza figli	27.771	19.558	35,82	53,79
Coppie con figli	26.677	59.269	80,37	12,78
Monogenitore	8.164	28.767	58,88	30,61
Altro	16.037	38.188	67,06	24,72
TOTALE				
Unipersonale	91.105	25.917	58,14	32,35
Coppie senza figli	31.568	51.738	36,05	53,90
Coppie con figli	30.805	68.536	81,06	12,22
Monogenitore	17.130	38.873	68,59	21,51
Altro	18.962	51.282	67,32	24,48
<b>Firenze</b>				
TIPOLOGIA FAMILIARE	N famiglie	Reddito familiare medio	% reddito da lavoro	% reddito da pensione
SOLO 1 PERCETTORE IN FAMIGLIA				
Unipersonale	75.783	25.844	55,02	33,56
Coppie senza figli	4.678	29.534	30,33	62,59
Coppie con figli	3.954	37.129	88,88	6,32
Monogenitore	8.334	28.608	81,45	8,97
Altro	3.746	20.945	79,67	14,54
DUE O PIÙ PERCETTORI IN FAMIGLIA				
Unipersonale	-	-	-	-
Coppie senza figli	21.834	52.375	34,21	54,42
Coppie con figli	27.866	68.132	78,79	14,41
Monogenitore	8.741	46.239	59,17	30,03
Altro	17.555	54.472	68,91	22,58
TOTALE				
Unipersonale	75.783	25.844	55,02	33,56
Coppie senza figli	26.512	48.345	33,79	55,30
Coppie con figli	31.820	64.279	79,52	13,83
Monogenitore	17.075	37.634	67,43	22,22
Altro	21.301	48.576	69,73	21,97

#Disagio

Grafico 4.2 - Composizione media del reddito per classe di età e condizione di disabilità. Anno 2011





**Tavola 4.4 - Numero di famiglie e reddito mediano delle famiglie con almeno un percettore di reddito, con o senza membri con disabilità, per tipologia familiare. Anno 2011. (Reddito mediano in euro)****MODENA**

TIPOLOGIA FAMILIARE	Famiglie con almeno un percettore di reddito				N. Famiglie senza percettori con almeno un disabile
	Famiglie con almeno un disabile		Famiglie senza disabili		
	N.	Reddito mediano	N.	Reddito mediano	
Unipersonale	2.829	19.626	27.213	20.626	0
Coppie senza figli	2.466	38.811	11.899	32.902	2
Coppie con figli	1.622	54.530	16.030	49.164	1
Monogenitore	772	30.548	5.896	35.732	0
Altro	1.870	36.811	10.265	47.963	0
<b>Totale</b>	<b>9.559</b>	<b>30.872</b>	<b>71.303</b>	<b>31.829</b>	<b>3</b>

**BOLOGNA**

TIPOLOGIA FAMILIARE	Famiglie con almeno un percettore di reddito				N. Famiglie senza percettori con almeno un disabile
	Famiglie con almeno un disabile		Famiglie senza disabili		
	N.	Reddito mediano	N.	Reddito mediano	
Unipersonale	7.433	20.881	83.672	20.545	0
Coppie senza figli	4.452	33.257	27.116	39.864	0
Coppie con figli	2.464	49.276	28.341	54.585	0
Monogenitore	1.858	34.000	15.272	29.428	2
Altro	3.090	44.390	15.872	39.750	0
<b>Totale</b>	<b>19.297</b>	<b>29.904</b>	<b>170.273</b>	<b>28.498</b>	<b>2</b>

**FIRENZE**

TIPOLOGIA FAMILIARE	Famiglie con almeno un percettore di reddito				N. Famiglie senza percettori con almeno un disabile
	Famiglie con almeno un disabile		Famiglie senza disabili		
	N.	Reddito mediano	N.	Reddito mediano	
Unipersonale	6.369	20.379	69.414	19.636	0
Coppie senza figli	3.641	31.550	22.871	37.920	0
Coppie con figli	2.326	46.824	29.494	52.441	0
Monogenitore	1.663	32.516	15.412	29.412	0
Altro	2.803	44.509	18.498	35.609	0
<b>Totale</b>	<b>16.802</b>	<b>29.055</b>	<b>155.689</b>	<b>28.896</b>	<b>0</b>

## #Istruzione e lavoro

**Tavola 4.5 - Numero di famiglie con nessuna, una, due o più persone in età attiva (15-64) che non studiano, non lavorano senza reddito, per numero di componenti. Anno 2011****MODENA**

NUMERO DI COMPONENTI	N. famiglie					
	Tutte le persone in età attiva (15-64) che studiano e/o lavorano e/o percepiscono reddito		Una persona in età attiva (15-64) che non studia, non lavora senza reddito		Due o più persone in età attiva (15-64) che non studiano, non lavorano senza reddito	
	V.a.	Val. %	V.a.	Val. %	V.a.	Val. %
1	30.605	93,1	2.276	6,9	-	-
2	21.426	90,7	2.008	8,5	195	0,8
3	11.628	80,5	2.492	17,3	321	2,2
4	7.266	74,2	2.073	21,2	456	4,7
5 e più	2.084	53,5	1.148	29,5	666	17,1
<b>Totale</b>	<b>73.009</b>	<b>86,3</b>	<b>9.997</b>	<b>11,8</b>	<b>1.638</b>	<b>1,9</b>

**BOLOGNA**

NUMERO DI COMPONENTI	N. famiglie					
	Tutte le persone in età attiva (15-64) che studiano e/o lavorano e/o percepiscono reddito		Una persona in età attiva (15-64) che non studia, non lavora senza reddito		Due o più persone in età attiva (15-64) che non studiano, non lavorano senza reddito	
	V.a.	Val. %	V.a.	Val. %	V.a.	Val. %
1	93.590	90,8	9.527	9,2	-	-
2	48.215	89,5	5.119	9,5	562	1,0
3	21.632	77,8	5.281	19,0	898	3,2
4	10.691	72,2	3.294	22,2	829	5,6
5 e più	2.427	52,0	1.390	29,8	849	18,2
<b>Totale</b>	<b>176.555</b>	<b>86,4</b>	<b>24.611</b>	<b>12,0</b>	<b>3.138</b>	<b>1,5</b>

**FIRENZE**

NUMERO DI COMPONENTI	N. famiglie					
	Tutte le persone in età attiva (15-64) che studiano e/o lavorano e/o percepiscono reddito		Una persona in età attiva (15-64) che non studia, non lavora senza reddito		Due o più persone in età attiva (15-64) che non studiano, non lavorano senza reddito	
	V.a.	Val. %	V.a.	Val. %	V.a.	Val. %
1	77.791	90,8	7.865	9,2	-	-
2	41.240	87,8	5.142	10,9	587	1,2
3	21.775	75,6	5.995	20,8	1.037	3,6
4	12.468	69,5	4.260	23,8	1.207	6,7
5 e più	2.987	50,6	1.743	29,6	1.168	19,8
<b>Totale</b>	<b>156.261</b>	<b>84,3</b>	<b>25.005</b>	<b>13,5</b>	<b>3.999</b>	<b>2,2</b>

#### 4.4 Sviluppi futuri

La sperimentazione realizzata nel corso del 2013 ha permesso di costruire delle collezioni di microdati a partire dall'integrazione di fonti amministrative, che consentono di produrre statistiche sulle famiglie relativamente a specifiche dimensioni informative. Tali basi dati possono essere interrogate in maniera flessibile, a seconda delle esigenze conoscitive dei potenziali utenti.

Si rileva, tuttavia, l'esigenza di ampliare ulteriormente il *framework* degli archivi amministrativi al fine di rendere disponibili altre informazioni su sotto-popolazioni di interesse e di affinare il più possibile il calcolo del reddito familiare. In particolare, si manifesta la necessità di disporre del titolo di studio di ogni componente della famiglia in quanto esso costituisce una informazione indispensabile per la dimensione "Istruzione" ai fini dell'analisi statistica sulla vulnerabilità sociale. Infatti, la relazione strettissima tra istruzione formale e livelli retributivi, probabilità di disoccupazione e tipo di partecipazione al mercato del lavoro è stata evidenziata da numerose ricerche.<sup>57</sup> Analoga esigenza si deve esprimere in merito alla futura disponibilità della variabile titolo di godimento dell'abitazione della famiglia ai fini della copertura della dimensione "Condizione abitativa". In relazione alla dimensione economica, si rileva la necessità di approfondire i contenuti informativi delle fonti già utilizzate e di acquisirne di nuove per perfezionare il calcolo del reddito. Di particolare interesse è la fonte Inps "Gestione fiscale delle prestazioni non pensionistiche" utile sia per la corretta classificazione delle voci del reddito familiare, sia per la derivazione di informazioni essenziali afferenti alle tematiche del mercato del lavoro e del disagio.

Ulteriore sviluppo del sottoprogetto riguarda l'applicazione di metodologie per la classificazione delle famiglie. Infatti, oltre ai diversi utilizzi che gli utenti possono fare dei microdati, ad esempio valutazione delle politiche sociali, si manifesta l'interesse di disporre di un'informazione relativa alla condizione socio-economica di ogni famiglia che sia allo stesso tempo riassuntiva e trasversale alle varie dimensioni. Ciò si traduce nella produzione di indicatori di sintesi per ottenere un *ranking* di tutte le famiglie di un dato territorio. Una delle ipotesi di lavoro per giungere ad una classificazione delle famiglie è costituita dall'impiego delle tecniche del Data Mining (DM) sui microdati del sotto-progetto. Un possibile ambito di approfondimento, oltre alla classificazione, riguarda l'applicazione di tali tecniche ai dati che rappresentano il frutto dell'integrazione tra famiglie fiscali e famiglie anagrafiche.

Per quanto riguarda l'integrazione tra famiglie fiscali e famiglie anagrafiche è stato condotto un primo test, utilizzando i dati delle LAC e quelli desumibili dal quadro dei familiari a carico (delle dichiarazioni Unico Persone Fisiche e 730), relativamente a circa 60 mila individui (si tratta degli individui sui cui viene svolta l'indagine Eu-Silc).<sup>58</sup> Questo test si pone lo scopo di valutare il valore informativo aggiunto derivante dalla mancata completa sovrapposizione delle famiglie anagrafiche con le famiglie fiscali. Infatti, se le LAC consentono di determinare le "famiglie anagrafiche" e le dichiarazioni reddituali "le famiglie fiscali", l'integrazione di questi due bacini informativi può fornire interessanti spunti per individuare situazioni in cui le dichiarazioni fiscali creano un legame tra due o più famiglie anagrafiche, riconoscendo i casi in cui i familiari a carico non risultano residenti con il dichiarante. I risultati del test indicano che è possibile far emergere relazioni tra persone non conviventi e arricchire la base anagrafica articolando maggiormente l'informazione della tipologia familiare.

---

<sup>57</sup> Checchi e Flabb (2006).

<sup>58</sup> Ceccarelli et al. (2008).

## 5. Conclusioni

Lo sviluppo sperimentale, realizzato nel corso del 2013, ha avuto come principale obiettivo quello di valutare l'effettiva potenzialità del progetto Archimede, con particolare riferimento all'identificazione di nuovi ambiti informativi direttamente ricavabili dai dati amministrativi integrati nel SIM. Le principali potenzialità dei tre sottoprogetti sperimentali realizzati possono essere così riassunte:

- Esiste la possibilità di caratterizzare a più livelli, fenomeni complessi: questo permette di darne letture integrative e molto specifiche rispetto a quelle tradizionali.
- Esiste la possibilità di utilizzare le basi di dati amministrativi integrati per sostituire alcune informazioni raccolte ogni 10 anni dai Censimenti, e renderle disponibili con cadenza annuale. Il riferimento è in particolare alla tematica sulla mobilità e alla identificazione e caratterizzazione delle popolazioni che usano un territorio, sviluppata dal sotto progetto sulla popolazione insistente.
- È possibile indagare a più livelli, con l'obiettivo di darne letture differenti e molto specifiche, fenomeni complessi come la precarietà lavorativa, ambito tematico del secondo sottoprogetto. In questo senso, l'approccio seguito non considera in maniera esclusiva le tipologie di lavoratori in possesso di contratti atipici (criterio in genere utilizzato per identificare i "precari") ma integra altre informazioni disponibili (sulla CIG, sulla mono/pluri committenza, sul reddito, ecc.) permettendo di caratterizzare questa tipologia di lavoratori in maniera alternativa e di farne una lettura più attenta.
- Esiste la possibilità, come dimostrato nel terzo sottoprogetto, di associare numerose informazioni alle famiglie in modo da evidenziare una carenza di risorse in una o più dimensioni fondamentali (lavoro, reddito, formazione, disagio, ecc.). Questo approccio permette da un lato una valutazione complessiva delle condizioni socio-economiche delle famiglie, dall'altro consente di studiare scenari di rischio più complessi rispetto a quelli che utilizzano, ad esempio, la sola dimensione reddituale.
- Infine, la possibilità di ottenere basi di microdati per intere popolazioni e/o sottopopolazioni, rende possibile analisi per ambiti territoriali ristretti, anche sub-comunali, con frequenza annuale (in alcuni casi anche mensile, come per il progetto precarietà lavorativa). Si superano in tal modo sia i vincoli territoriali delle classiche indagini campionarie, che forniscono stime solo per determinati domini, sia i limiti temporali di quelle censuarie.

Tuttavia, resta ancora da valutare una serie di aspetti fondamentali. Si accennano di seguito i più rilevanti.

Il primo è sicuramente quello connesso alla "valutazione dell'utilità conoscitiva" del progetto in generale e delle basi informative prodotte in particolare. Dal momento che l'obiettivo del progetto è quello di produrre collezioni di microdati utili "*alla programmazione territoriale e settoriale, alla valutazione delle politiche pubbliche a livello nazionale, regionale e locale*", risulta evidente come una valutazione debba essere data proprio dai potenziali utenti. Per questo motivo l'Istat ha individuato come compartecipanti alcuni uffici di statistica di Regioni, Province e Comuni, i quali hanno il compito di valutare le basi di microdati prodotte dall'Istat con riferimento ai seguenti aspetti generali:

- a) completezza, congruità e utilità del contenuto informativo rispetto alle esigenze conoscitive di specifici fenomeni oggetto di studio connessi alle finalità istituzionali dell'ente di appartenenza dell'Ufficio di Statistica;
- b) vantaggi derivanti all'Ufficio di statistica in termini di risparmio di costi e di miglioramento della qualità dell'informazione statistica prodotta.

I soggetti compartecipanti, in accordo con l'Istat, hanno individuato specifici fenomeni di interesse (ad es. mobilità territoriale, autonomia familiare, dinamiche occupazione nel territorio, marginalità sociale, ecc.) su cui effettuare specifici test. I risultati sono previsti per i primi mesi del 2015.

Il secondo aspetto centrale è quello del *timing* dei dati resi disponibili. Risulta evidente come la programmazione settoriale e territoriale e la valutazione delle politiche abbiano necessità di infor-

mazioni sufficientemente vicine in termini temporali *ai* fenomeni da programmare e *dai* fenomeni da valutare. Questo aspetto è chiaramente connesso sia all'oggettiva, rispetto alle specifiche finalità, disponibilità delle fonti amministrative sia alla più generale politica dell'Istituto (valutazione delle priorità e delle risorse disponibili) nell'acquisizione e integrazione delle fonti. Detto ciò, un altro punto da valutare è relativo al *trade off* fra la tempestività delle informazioni e la loro qualità, in particolar modo in riferimento alla possibilità di acquisire dati amministrativi con copertura parziale e di qualità non eccellente, ma in tempi relativamente brevi.<sup>59</sup>

Infine, resta da dirimere l'aspetto connesso alla coerenza fra gli *output* prodotti dal progetto e quelli, simili, derivati da altri processi statistici. In questo frangente, si ritiene, in primo luogo, indispensabile predisporre una documentazione dettagliata sulla produzione degli *output* del progetto Archimede. In secondo luogo è sicuramente necessario documentare e, ove possibile, quantificare le differenze (concettuali, di processo, ecc..) che sussistono fra i risultati dei sottoprogetti e quelli derivanti dalle indagini condotte dall'Istituto, in particolare l'indagine sulle forze di lavoro (che rappresenta il benchmark ideale per il sottoprogetto relativo alla precarietà lavorativa) e l'indagine Eu-Silc (da utilizzare come riferimento per verificare la coerenza delle informazioni reddituali).

Tuttavia, non è detto che queste informazioni siano adeguate o sufficienti. Un approccio nuovo che l'Istituto sta sperimentando è quello della definizione di Sistemi Informativi specifici che integrano informazioni amministrative (riferite ad intere popolazioni) con dati campionari da indagine. Si tratta sostanzialmente di sistemi intermedi che partendo da un trattamento statistico delle fonti amministrative, associano alle unità o ad alcune caratteristiche una probabilità stimata da modello utilizzando i risultati delle indagini campionarie. Questo approccio è in fase di studio con riferimento sia ai dati su lavoro e occupazione sia a quelli sui redditi degli individui. Questi sistemi, seppur non totalmente risolutivi delle incoerenze, possono rappresentare, strumenti che permettono la riduzione (e la corretta quantificazione) delle discrepanze che si possono avere tra i dati derivabili dalle fonti amministrative e le stime prodotte attraverso indagini.

---

<sup>59</sup> Questo aspetto è parte dell'attività previste dal GdL ArcoIaio istituito dall'Istat (delibera DGEN N°65/14 del 10/7/2014), che ha il compito di fornire metodologie per misurare la qualità statistica dei processi produttivi che utilizzano dati amministrativi.

## Riferimenti bibliografici

- Agenzie delle Entrate. 2013. *Statistiche Catastali 2012*.  
[http://www.agenziaentrate.gov.it/mt/osservatorio/Tabelle%20statistiche/StatisticheCatastali2012\\_29102013.pdf](http://www.agenziaentrate.gov.it/mt/osservatorio/Tabelle%20statistiche/StatisticheCatastali2012_29102013.pdf) (1 ottobre 2014).
- Ceccarelli, C., Di Marco, M., Rinaldelli, C. eds. 2008. *L'indagine europea sui redditi e le condizioni di vita delle famiglie (Eu-Silc)*. Roma.
- Checchi, D., Flabbi, L. 2006. *Mobilità intergenerazionale e decisioni scolastiche in Italia*. In Ballarino, G., Checchi, D. eds. *Scelte individuali e vincoli strutturali. Sistema scolastico e disuguaglianza sociale*. Bologna: il Mulino.
- Ambroselli, S., Di Bella, G. 2024. *Towards a more efficient system of administrative data management and quality evaluation to support statistics production in Istat*. Eurostat, Q2014, Wien.
- European Commission. 2000. *Declaration by the Council and the Commission on the European Community's development policy*.
- European Commission. 2012. *Study on Precarious work and social rights*. Working Lives Research Institute. Faculty of Social Sciences and Humanities - London Metropolitan University.
- Istat (a cura di) Gazzelloni, S. 2006. *La rilevazione sulle forze di lavoro: contenuti, metodologie, organizzazione*. Roma.
- Istat. 2010. *Rilevazione sulle forze di lavoro Media 2009*.
- Istat. 2011(a), *La metodologia di stima dei redditi lordi nell'indagine Eu-Silc*. Metodi e norme n.49.
- Istat. 2011(b). *I redditi delle famiglie con stranieri Anni 2008-2009*. Statistiche report.
- Istat. 2012. *Statistiche della previdenza e dell'assistenza sociale - I beneficiari delle prestazioni pensionistiche*.
- Istat. 2013. *La struttura informativa sull'occupazione di Asia per il CIS 2011*.
- Istat, Cnel. 2013. *BES 2013 Il benessere equo e sostenibile in Italia*.
- Istat, Servizio REG. 2013. *La struttura informativa sull'occupazione di ASIA per il CIS2011* (documento interno).
- Jenks, G. 1977. *Optimal Data Classification for Choropleth Maps*.
- Mandrone, E. 2008. *Quando la flessibilità diviene precarietà: una stima sezionale e longitudinale*. Studi Isfol 2008/6.
- Mandrone, E., Radicchia, D. 2012. *Indagine Plus: il mondo del lavoro tra forma e sostanza*. I libri del Fondo Sociale Europeo, n.167.
- Oecd. 2001. *The DAC Guidelines on Poverty Reduction*. Paris.
- Tsakoglou, P., Papadopoulos, F. 2002. *Aggregate level and determining factors of social exclusion in twelve European countries*. Journal of european social policy, 12(3).
- United Nations. 2011. *Canberra Group Handbook on Household Income Statistics*.

## Informazioni per gli autori

La collana è aperta ad autori dell'Istat e del Sistema statistico nazionale, e ad altri studiosi che abbiano partecipato ad attività promosse dal Sistan (convegni, seminari, gruppi di lavoro, ecc.). Da gennaio 2011 essa sostituirà Documenti Istat e Contributi Istat.

Coloro che desiderano pubblicare sulla nuova collana dovranno sottoporre il proprio contributo alla redazione degli Istat Working Papers inviandolo per posta elettronica all'indirizzo [iwp@istat.it](mailto:iwp@istat.it). Il saggio deve essere redatto seguendo gli standard editoriali previsti, corredato di un sommario in italiano e in inglese; deve, altresì, essere accompagnato da una dichiarazione di paternità dell'opera. 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*.

Per gli autori Istat, la sottomissione dei lavori deve essere accompagnata da una mail del proprio dirigente di Servizio/Struttura, che ne assicura la presa visione. Per gli autori degli altri enti del Sistan la trasmissione avviene attraverso il responsabile dell'ufficio di statistica, che ne prende visione. Per tutti gli altri autori, esterni all'Istat e al Sistan, non è necessaria alcuna presa visione. Tutti i lavori saranno sottoposti al Comitato di redazione, che valuterà la significatività del lavoro per il progresso dell'attività statistica istituzionale. La pubblicazione sarà disponibile su formato digitale e sarà consultabile on line.

Gli articoli pubblicati impegnano esclusivamente gli autori, le opinioni espresse non implicano alcuna responsabilità da parte dell'Istat. Salvo diversa indicazione la riproduzione è libera, a condizione che venga citata la fonte.