



DATI AMMINISTRATIVI, METODI E STATISTICHE PER LE POLITICHE TERRITORIALI

A CURA DI
FLAVIO VERRECCHIA



FrancoAngeli

OPEN  ACCESS



Il presente volume è pubblicato in open access, ossia il file dell'intero lavoro è liberamente scaricabile dalla piattaforma **FrancoAngeli Open Access** (<http://bit.ly/francoangeli-oa>).

FrancoAngeli Open Access è la piattaforma per pubblicare articoli e monografie, rispettando gli standard etici e qualitativi e la messa a disposizione dei contenuti ad accesso aperto. Oltre a garantire il deposito nei maggiori archivi e repository internazionali OA, la sua integrazione con tutto il ricco catalogo di riviste e collane FrancoAngeli massimizza la visibilità, favorisce facilità di ricerca per l'utente e possibilità di impatto per l'autore.

Per saperne di più:

http://www.francoangeli.it/come_publicare/publicare_19.asp

I lettori che desiderano informarsi sui libri e le riviste da noi pubblicati possono consultare il nostro sito Internet: www.francoangeli.it e iscriversi nella home page al servizio "Informatemi" per ricevere via e-mail le segnalazioni delle novità.

DATI AMMINISTRATIVI, METODI E STATISTICHE PER LE POLITICHE TERRITORIALI

A CURA DI
FLAVIO VERRECCHIA

FrancoAngeli
OPEN  ACCESS

Si ringrazia il Dipartimento di Economia, Management e Metodi Quantitativi dell'Università degli Studi di Milano per il supporto dato a questo progetto nell'ambito del programma quinquennale 2018-2022 "Dipartimenti di eccellenza" finanziato dal Ministero dell'Istruzione, dell'Università e della Ricerca al fine di incentivare l'attività dei Dipartimenti delle Università statali che si caratterizzano per l'eccellenza nella qualità della ricerca e nella progettualità scientifica, organizzativa e didattica (legge 11 dicembre 2016, n. 232 - Art. 1, commi 314-338).

Copyright © 2019 by FrancoAngeli s.r.l., Milano, Italy.

Pubblicato con licenza *Creative Commons Attribuzione-Non Commerciale-Non opere derivate 4.0 Internazionale* (CC-BY-NC-ND 4.0)

L'opera, comprese tutte le sue parti, è tutelata dalla legge sul diritto d'autore. L'Utente nel momento in cui effettua il download dell'opera accetta tutte le condizioni della licenza d'uso dell'opera previste e comunicate sul sito

<https://creativecommons.org/licenses/by-nc-nd/4.0/deed.it>

INDICE

| | | |
|--|------|----|
| PREFAZIONE | pag. | 7 |
| <i>di G. Manzi</i> | » | 9 |
| INTRODUZIONE AL VOLUME | | |
| <i>di G. Garofalo</i> | | |
| PARTE PRIMA – <i>Dati e metadati</i> | | |
| Introduzione | » | 11 |
| <i>di S. Figini e A. Venturi</i> | | |
| Capitolo 1 – Microdati integrati economici e demosociali | » | 13 |
| <i>di S. Casacci e D. Ercolani</i> | | |
| Capitolo 2 – Metadati per capire i dati | » | 21 |
| <i>di D. Ferrazza</i> | | |
| Capitolo 3 – Qualità dei dati amministrativi | » | 29 |
| <i>di M. E. Comune e L. C. M. Viviano</i> | | |
| PARTE SECONDA – <i>Metodi</i> | | |
| Introduzione | » | 37 |
| <i>di P.M. Chiodini</i> | | |
| Capitolo 4 – Gestire i dati amministrativi | » | 39 |
| <i>di S. Ballabio, D. Ferrazza e F. Verrecchia</i> | | |
| Capitolo 5 – Costruire gli indici sintetici | » | 47 |
| <i>di M. Mazziotta</i> | | |
| Capitolo 6 – Rappresentare le informazioni territoriali | » | 53 |
| <i>del GdL Istat - SPoT</i> | | |
| PARTE TERZA – <i>Applicazioni su dati amministrativi negli Enti locali</i> | | |
| Introduzione | » | 63 |
| <i>di A. Lentini</i> | | |
| Capitolo 7 – L’esperienza del Comune di Brescia per le politiche locali | » | 65 |
| <i>di M. Trentini</i> | | |
| Capitolo 8 – L’esperienza della Regione Lombardia per la previsione della domanda di lavoro e la definizione dell’offerta formativa | » | 75 |
| <i>di F. Rappelli e V. Ricciari</i> | | |

PARTE QUARTA – *Applicazioni sperimentali su dati amministrativi*

| | | |
|--|------|-----|
| Introduzione | pag. | 83 |
| <i>di M. Filandri</i> | | |
| Capitolo 9 – <i>Working poor</i> nel Nord-Ovest | » | 85 |
| <i>di S. Cariello e M. Carbonara</i> | | |
| Capitolo 10 – Proposta di un indice di vulnerabilità socioeconomica | » | 93 |
| <i>di R. Succi, F. Vannoni e A. Vitalini</i> | | |
| Capitolo 11 – Redditi delle famiglie e condizioni di fragilità in Lombardia | » | 101 |
| <i>di S. Casacci</i> | | |
| Capitolo 12 – Disparità di genere nel mercato del lavoro in Lombardia | » | 109 |
| <i>di A. L. Carulli, L. Sicuro e D. Tucci</i> | | |
| Capitolo 13 – Neet e Quasi-Neet in Lombardia | » | 117 |
| <i>del GdL Istat - SPoT Lombardia</i> | | |
| Capitolo 14 – Non profit in Lombardia: una risposta resiliente? | » | 123 |
| <i>di S. Della Queva, M. Nicosia e S. Stoppiello</i> | | |
| | | |
| CONCLUSIONI E PROSPETTIVE FUTURE | » | 131 |
| <i>di F. Verrecchia</i> | | |
| | | |
| APPENDICI | | |
| Appendice 1 – Redditi | » | 133 |
| Appendice 2 – Comuni Frontalieri | » | 135 |
| Appendice 3 – <i>Missing data</i> | » | 137 |
| | | |
| RIFERIMENTI BIBLIOGRAFICI | » | 139 |
| | | |
| AUTORI | » | 147 |
| | | |
| RINGRAZIAMENTI | » | 149 |
| | | |
| WORD CLOUDS DEI CAPITOLI | » | 151 |

PREFAZIONE

di Giancarlo Manzi

“Data! data! data!” he cried impatiently. “I can’t make bricks without clay.” Questa esclamazione – attribuita a Sherlock Holmes nel racconto “L’avventura dei faggi rossi” di Conan Doyle, quando le supposizioni di Watson non portavano a nulla di dimostrabile per risolvere il caso del mistero dell’ala proibita della casa del signor Rucastle, in una delle più famose avventure del *detective* britannico – non è mai stata così attuale come in quest’era della cosiddetta *rivoluzione dei big data*.

Malgrado non esista una separazione rigorosa tra *big data* e altri dati, il termine *big data* indica genericamente una raccolta di dati così estesa da richiedere tecnologie e metodi analitici specifici per l’estrazione di valore o conoscenza.

L’apporto informativo dei *big data* ha avuto un impatto enorme in ogni campo. Se si considera, ad esempio, l’unità di misura più comune per le dimensioni trattate in ambito di analisi dei dati, in pochi anni si è passati dall’utilizzo del gigabyte (10^9 byte) allo zettabyte (10^{21} byte) e più recentemente allo yottabyte (10^{24} byte). Le dimensioni dei *data center* di Google arrivano a comprendere estensioni ormai paragonabili a decine di campi di calcio, così come il loro consumo di energia richiede impianti di raffreddamento equiparabili a quelli delle centrali nucleari.

Ma non è solo relativamente alle dimensioni delle basi di dati da gestire che le tecnologie informatiche applicate ai dati hanno fatto passi da gigante. In generale, infatti, oltre al “volume” l’evoluzione ha riguardato anche la “velocità”, la “varietà”, il “valore” e la “veridicità” dei dati, vale a dire tutte le “5 V” che molti studiosi utilizzano per sintetizzare le “componenti” dei *big data*. A queste vanno aggiungendosi altre componenti dell’“ecosistema” dei *big data*, come il “*real-time analytics*”, il “*data linkage*” (Demchenko et al., 2014) e, nel tempo, altre ancora.

La crescente facilità di accesso, di gestione e l’economicità dei *big data* non è priva di controindicazioni. Ci sono infatti anche dei rischi legati alla loro diffusione. Ad esempio, se da un lato si tende sempre più a cercare in essi la soluzione informativa unica e risolutiva rispetto alle esigenze conoscitive di una società globale immersa in un processo evolutivo sempre più veloce e coinvolgente, dall’altro, spesso si dimentica di analizzare le forzature metodologiche sottostanti il loro utilizzo. In tal senso, se si considera il confronto tra una fonte di *big data* come i *social network* e l’indagine d’opinione tradizionale, le informazioni provenienti dai *social network* sono più tempestive (la V della velocità), più veritiere rispetto al *sentiment* reale (la V della veridicità) e più commerciabili (la V del valore). Tuttavia, al pari, non si può tralasciare di considerare che le analisi basate sui *big data* hanno problemi metodologici non trascurabili nell’ambito dell’intero processo informativo: dalla corretta interpretazione testuale nel caso dei *tweets*, alla non rappresentatività di intere sezioni della popolazione, all’incertezza nella valutazione della variabilità delle stime, ecc. Il *mainstream*, a favore dei *big data* e degli *analytics*, tende a non considerare queste problematiche che invece non possono essere trascurate, soprattutto dagli enti certificatori dell’informazione come gli Istituti nazionali di statistica.

Pertanto, nell'ambito della Statistica ufficiale, la sfida consiste nel considerare fonti di informazione quantitativa sempre nuove ma isolandone i difetti e valorizzandone i pregi. Oggi in Italia l'Istat si è dotato di una infrastruttura orientata allo studio e alle analisi basate sui *big data*, tanto che anche il censimento della popolazione – diventato permanente – ha visto l'adeguamento, coerentemente ai tempi che cambiano, delle procedure di rilevazione dei dati. Infatti, le informazioni su cittadini e famiglie sono fortemente basate sulle anagrafi dei comuni e su altre fonti amministrative che possono essere considerate vere proprie fonti di informazione di tipo *big data*.

Ecco dunque che anche strumenti di indagine statistica tradizionali come l'indagine campionaria, se adeguati e integrati al mondo dei *big data*, diventano anche più importanti rispetto al passato perché, possono certificare e validare l'informazione proveniente dai *big data*. Al giorno d'oggi, infatti, conviene integrare le fonti di informazione piuttosto che metterle in competizione se si vogliono ottenere i migliori risultati in termini di bontà dell'informazione statistica.

Tra le fonti di informazione che sono maggiormente sfruttate per una conoscenza più completa dei fenomeni collettivi e che devono essere trattate come *big data* ci sono appunto le fonti di dati amministrativi. L'Istat, per alcuni fenomeni, ha tradizionalmente utilizzato il dato amministrativo quale fonte statistica importantissima da integrare con tutte le altre. Tuttavia, le importanti potenzialità dello sfruttamento del dato amministrativo con finalità conoscitive sono state finora limitate per motivi sia tecnici sia di integrazione e di coordinamento fra i vari Enti titolari delle basi dati amministrative.

Questo volume raccoglie alcune sperimentazioni nate nell'ambito del sistema formato da amministrazioni pubbliche ed enti privati che produce tutte le statistiche ufficiali in Italia (Sistan), volte all'integrazione dei dati amministrativi – oltre che rispetto alle altre fonti di informazione – e al loro utilizzo per lo studio dei fenomeni finalizzato al governo dei territori. L'integrazione è importante soprattutto nel processo di armonizzazione delle fonti statistiche territoriali. La tempestività e la correttezza dell'informazione statistica a livello territoriale di granularità sub-provinciale assume importanza crescente soprattutto quando il dato elementare può essere aggregato e integrato così facilmente come oggi le tecniche informatiche consentono.

Migliorare le tecniche di integrazione delle fonti di dati, qualsiasi esse siano, aumentandone l'efficienza – anche riguardo alla riduzione dei costi – è diventato dunque estremamente importante. Allo stesso tempo, i dati amministrativi hanno tutto il potenziale per diventare una fonte di informazione rilevante per la statistica ufficiale. È necessario allora aumentare le sinergie tra di essi, captarne le potenzialità di *linkage* e di aggregazione e sfruttare le moderne tecnologie dell'*information technology*. È quello che il Sistan sta cercando di fare in questi ultimi anni e questo volume è uno dei risultati di questa prospettiva.

INTRODUZIONE AL VOLUME

di Giuseppe Garofalo

L'Istituto nazionale di statistica ha intrapreso un percorso di profonda revisione nei processi di produzione delle statistiche pubbliche. Tale percorso ha nella costruzione di un Sistema integrato dei registri (SIR) il suo punto nevralgico. Tale sistema si basa sullo sfruttamento massivo, organizzato ed integrato di informazioni acquisite per fini *non* statistici (dati amministrativi, ma non solo, anche i cosiddetti *big data*) e indagini statistiche, di tipo campionario, che hanno la funzione di integrare le informazioni mancanti e di correggere (per sopra e sotto copertura e per *miss classification*) le informazioni *esterne* acquisite.

Su questa idea si regge il processo dei censimenti permanenti.

Risulta evidente la complessità del percorso in atto, per altro già operativo per la parte delle statistiche sulle imprese, per il quale non solo è necessario un grande sforzo organizzativo e tecnologico ma anche, e soprattutto, lo sviluppo di un sistema organico di meta informazioni¹, la realizzazione di metodologie statistiche specifiche, un sistema di valutazione della qualità statistica a livello di microdato. Allo stesso tempo, risulta chiaro come l'approccio in fase di realizzazione permetterà di risolvere, con costi sicuramente contenuti, due aspetti che la statistica pubblica ha sempre avuto difficoltà ad affrontare. Da un lato la possibilità di avere una ampio raggio di informazioni – su lavoro, istruzione, reddito, relazioni familiari, ecc. – in numero superiore a quello dei classici censimenti, per ambiti territoriali di piccola dimensione anche sub-comunali. Dall'altro la possibilità di permettere studi e valutazioni sui fenomeni sociali ed economici a partire da una analisi multidimensionale (Cfr. "Introduzione alla parte III e IV"), permettendo di superare, al contempo, la tipica e classica separazione della statistica pubblica fra *famiglia-analisi sociale* e *impresa-analisi economica*.

È all'interno di questo quadro innovativo e in sviluppo che si basa l'esperienza del progetto ArchIMEDe, da cui sono partite la gran parte degli studi presentati in questo volume.

Il progetto ArchIMEDe (Cfr. Capitoli I e II) nasce nel 2013-2014 per rispondere alla domanda: la disponibilità nell'Istituto di una grande mole di dati amministrativi (del Miur, dell'Agenzia delle Entrate, dell'Inps, delle CCIAA, dei Comuni, ecc.) può essere *di per sé* utilizzata per la produzione di informazioni utili per una analisi territoriale di fenomeni sociali ed economici?

La risposta sembra essere positiva. Le basi dati di ArchIMEDe sono correntemente diffuse dall'Istat tramite la piattaforma Armida. A partire dal 2018 (con dati al 2015) 12 indicatori – su lavoro, mobilità, condizioni delle famiglie – saranno diffusi nell'ambito delle Statistiche Sperimentali dell'Istat² per gran parte dei comuni Italiani³, nell'ambito del progetto "Misure del benessere e della programmazione a livello comunale".

Gli stessi studi, effettuati da ricercatori Istat e di Enti Sistan, proposti in questa pubblicazione permettono di valutare le potenzialità informative che l'utilizzo di basi di dati come

¹ Per affrontare e risolvere tali problematiche si è scelto di adottare un approccio basato sull'uso di ontologie. Le ontologie sono modelli formali per la rappresentazione concettuale dei dati e, sulla base dei recenti avanzamenti tecnologici, sono esprimibili tramite linguaggi computazionali standard.

² Le statistiche sperimentali, diffuse a partire da marzo 2018 in un nuovo sito istituzionale, permettono di offrire i risultati delle sperimentazioni (su nuove fonti, nuove metodologie, nuovi indicatori) alla fruizione e alla valutazione degli utenti, chiarendo il loro status di statistiche sperimentali e non di statistiche ufficiali.

³ È in fase di valutazione il criterio di diffusione degli indicatori per i comuni di più piccola dimensione.

quelle di ArchIMEDe hanno in particolare per l'analisi e la programmazione territoriale, in particolare per la possibilità di calcolare indicatori (Cfr. "Indice di vulnerabilità socioeconomica" nel capitolo X) non possibili con altri tipi di informazioni, quali dati censuari e da indagini campionarie. Nello stesso tempo è necessario avere ben chiaro i *limiti* nell'utilizzo di queste informazioni e i *criteri di lettura* che di esse si possano fare.

Al netto di problemi connessi alla sottocopertura (Cfr. Capitoli III e VIII e Appendici 2 e 3), quali l'assenza delle liste anagrafiche di alcuni comuni, l'impossibilità di acquisire alcune informazioni (frontalieri), l'assenza di alcune fonti non ancora acquisite dall'Istat (Afam e formazione professionale nell'ambito dell'istruzione); tre sono gli elementi, evidenziati nei capitoli del volume, che bisogna tenere ben presenti quando si utilizzano tali tipologie di dati e li si comparano con quelli delle statistiche ufficiali prodotte dall'Istat.

Il primo è legato alla differenza di processo e di fonti utilizzate per il calcolo di statistiche. L'utilizzo di dati esaustivi vs indagini campionarie, a parità di altre caratteristiche, di per sé determina differenze molto rilevanti. Il secondo è connesso alla definizioni utilizzate. Le definizioni amministrative non sono quelle della statistica ufficiale e quindi i valori statistici determinati con definizioni differenti difficilmente possono essere comparate. Il caso più evidente è quello dei redditi (Cfr. Capitolo XI) in cui utilizzare il reddito lordo delle famiglie, invece di quello netto disponibile determina la perdita del fattore redistributivo che si ha con la tassazione. Infine differenze si hanno con riferimento al metodo di calcolo di alcuni aggregati, i metodi di calcolo risentono delle caratteristiche dei dati a disposizione (p.es. l'occupazione come media settimanale rispetto alla media mensile) e determinano il sovra o il sotto dimensionamento dei valori calcolabili⁴. Ma se i livelli assoluti di variabili o indicatori "comparabili" presentano differenze anche notevoli fra i dati desunti da ArchIMEDe e quelli delle indagini campionarie utilizzate per la statistica ufficiale, il *rank* ("graduatoria") per ambiti territoriali (ove comparabili, quindi regioni e solo in alcuni casi province) rimane sostanzialmente uguale (Bianchino e Garofalo, 2017). L'approccio utilizzato nell'ambito di questo volume e del Gruppo di lavoro SPoT, basato su indicatori e rappresentazioni grafiche (Cfr. Capitoli V e VI), ha permesso di sviluppare "*modalità di rappresentazione dei dati di fonte amministrativa che, pur mantenendo un elevato contenuto informativo, garantissero risultati coerenti con le statistiche ufficiali*"⁵.

Infine, nel Capitolo VII, Marco Trentini apre ad un interessante approccio, poco discusso nell'ambito degli Statistici Pubblici: la programmazione e valutazione delle politiche locali ha bisogno di informazioni desunte da dati amministrativi, perché l'intervento delle amministrazioni è essenzialmente amministrativo. I *warning* nell'interpretazione dei risultati che si ottengono utilizzando basi di dati amministrative integrate accanto al processo di sviluppo del SIR fanno assumere alle basi di dati di ArchIMEDe un carattere di *provvisorietà*. Si può affermare che in questa fase di passaggio ArchIMEDe svolge una funzione di supplenza. Una funzione che sarà garantita fino al completo sviluppo del SIR e che comunque permetterà di inglobare i micro dati *statistici* dei registri che sono progressivamente realizzati: a partire dal 2018 con i dati del Registro del lavoro, per proseguire nel 2019, e alla fine della prima edizione del censimento permanente, con quelli del Registro base degli individui e delle famiglie.

⁴ È questo il caso del calcolo dell'intensità lavorativa. In ArchIMEDe l'intensità lavorativa si configura come una misura di presenza annuale nelle fonti sull'occupazione analizzate; la quota annuale è computata come media dell'intensità lavorativa mensile; l'intensità mensile è calcolata sulla base delle settimane lavorate. Nel caso dell'indagine EU_SILC, l'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 e il numero totale di mesi teoricamente disponibili per attività lavorative. In EU_SILC non vi è una quantificazione della quota mensile lavorata e quindi tendenzialmente l'indagine produce valori superiori a quelli di ArchIMEDe.

⁵ Cfr. § 6.1.

INTRODUZIONE ALLA PARTE PRIMA

DATI E METADATI

di Silvia Figini e Alessandro Venturi

La statistica ufficiale a supporto delle scienze economiche e sociali si sta indirizzando verso un percorso innovativo di raccolta del dato sia a livello nazionale sia internazionale. Nuove metodologie e algoritmi sono stati studiati per integrare e combinare differenti forme di informazioni con le fonti tradizionali delle indagini e degli archivi amministrativi.

Il dato statistico nasce dalla lettura di un fenomeno reale di interesse, ed essendo rilevato da un'indagine o da un archivio amministrativo preesistente, è preventivamente definito e successivamente rilevato, generalmente insieme al processo produttivo che lo genera.

Le metodologie di *machine learning* e di analisi del dato consentono di integrare dati provenienti da fonti differenti: rilevazioni tradizionali, dati già in possesso della pubblica amministrazione e altre fonti innovative come i *big Data* con l'obiettivo di offrire in modo tempestivo informazioni di qualità.

I *big data* hanno caratteristiche differenti rispetto alle fonti dati tradizionali e richiedono metodi di raccolta e trattamento dei dati diversi da quelli comunemente usati, che si traducono in creazioni di infrastrutture e sviluppo di nuovi *software* di analisi.

L'Istat ha avviato diversi progetti sull'uso dei *big data*: *Persons and Places (Mobile Phone Data)* con lo scopo di integrare l'uso di dati della telefonia mobile nel processo statistico di stima di flussi di popolazione intercomunale; *Labour Market Estimation (Google Trends)* per previsioni sul *nowcasting*. Altri progetti avviati riguardano l'uso degli *Scanner data* per la stima dei prezzi al consumo; di tecniche di *web-scraping* per la stima dell'utilizzo di *Ict* da parte delle imprese e dei "Social media" (*Twitter, Facebook*) per la stima di indicatori di fiducia.

La sfida principale ancora oggi per l'uso del dato e delle nuove metodologie di analisi è quella di passare dalla sperimentazione alla produzione. Questo passaggio coinvolge diversi aspetti che spaziano dalla normativa sulla *privacy* alla necessità di dotarsi di nuove infrastrutture (metodologiche, tecnologiche, organizzative) nonché di nuove competenze.

La crescente disponibilità di dati e l'esigenza di disporre di informazioni di qualità, se da una parte si sta quindi rivelando determinante per soddisfare esigenze di conoscenza dei fenomeni sociali ed economici di una società sempre più complessa (e per rendere sostenibile il costo della raccolta dati), dall'altra comporta la non facile risoluzione di questioni oltre che sul fronte tecnico, anche su quello metodologico ed organizzativo. Considerando una generica popolazione statistica (quale ad esempio famiglie e/o singoli individui, imprese e/o istituzioni, etc.), il quadro generale per la produzione di informazioni statistiche si presenta sempre più spesso caratterizzato dalla coesistenza di fonti diverse di dati, originate dalla popolazione *target* (o ad essa correlate). Se da un lato sui dati statistici raccolti mediante rilevazioni tradizionali le metodologie e i processi sono da tempo consolidati e standardizzati a livello internazionale, l'introduzione dei *big Data* nella statistica ufficiale ha implicato la definizione e risoluzione di tre principali problematiche

che spaziano dal classificare e sistematizzare le fonti dati, al valutare come le caratteristiche specifiche dei *big Data* (tipicamente volume, varietà e velocità) hanno impatto e possono essere trattate in ambito statistico, al distribuire lungo il ciclo di vita dei dati statistici (raccolta, validazione, analisi e utilizzo) il loro trattamento.

Anche il concetto di Statistica inteso come disciplina scientifica che studia i metodi di raccolta, di organizzazione, di analisi, d'interpretazione e presentazione dei dati, è cambiato. Oggi si parla di *data science*, o *data-driven science*, come approccio metodologico multidisciplinare volto ad estrarre conoscenza dai dati e in quest'ambito, l'informatica, le normative per il trattamento informatico dei dati e i domini applicativi di interesse rivestono un ruolo cruciale.

L'estrazione di informazioni e conoscenza attraverso la scienza dei dati è possibile solo se l'insieme delle competenze statistiche, delle tecniche per il trattamento informatico dei dati e dell'informatica saranno usate non solo per finalità pratiche e commerciali ma soprattutto per finalità scientifiche e di ricerca.

Più in particolare, le tematiche considerate nella prima parte del volume riguardano sia i dati da fonte secondaria ed i relativi metadati, sia aspetti di interesse inerenti alla qualità degli archivi amministrativi su individui e famiglie. Grazie alle innovazioni *Ict* nella pubbliche amministrazioni, si è assistito nell'ultimo ventennio ad una proliferazione di archivi da fonte amministrativa di tipo *big data* con realizzazioni e applicazioni statistiche. Se da un lato possiamo dire consolidate le applicazioni su dati amministrativi relative alle imprese dall'altro lato diverso è il discorso per individui e famiglie. È evidente infatti come l'evoluzione delle capacità tecniche, ad esempio per il censimento delle imprese, abbia consentito di passare da una "indagine assistita da lista" – nel 2001 – al "registro assistito da indagine (i.e. ASIA-IULGI)" – nel 2011. La complessità delle fonti amministrative sugli individui è tale da poterle assimilare a fonti di tipo *big data*. Di contro, in pochi anni si è assistito ad una accelerazione dello sfruttamento di diverse nuove fonti amministrative: Uniemens e di Persone d'impresa (fonte CCIAA) per ASIA-occupazione; delle anagrafi comunali per il censimento della popolazione.

Proprio in questa scia, si inseriscono le recenti attività orientate allo sfruttamento a fini statistici dei dati amministrativi su individui e famiglie fortemente basate sull'integrazione di archivi. ArchIMEDe (Cfr. Capitolo I e II) rappresenta l'esperienza più evoluta nella direzione progettuale considerata e la cui sperimentazione a fini statistici deve considerare precisi requisiti di qualità (Cfr. Capitolo III) connessi agli obiettivi di studio dei fenomeni anche in declinazione territoriale.

Big data, dati amministrativi, metadati, open data sono di grande interesse per finalità istituzionali, imprenditoriali, e oggi giocano un ruolo fondamentale nelle applicazioni di *data science*. Queste devono saper affrontare in modo metodologicamente corretto problemi come eterogeneità, integrazione ed errori.

L'auspicio è che la ricerca consenta di produrre un ragionamento statistico più efficiente per derivare informazioni utili per prendere decisioni in base alle nuove tipologie di dati.

CAPITOLO I

MICRODATI INTEGRATI ECONOMICI E DEMOSOCIALI¹

di Sara Casacci e Dario Ercolani

1.1 Premessa

Da diversi anni l'Istat promuove l'uso intensivo delle fonti amministrative sia per contenere i costi legati alle attività di rilevazione diretta, sia per soddisfare la crescente richiesta di informazione statistica nuova e di maggior dettaglio territoriale. In tale contesto si colloca il progetto ArchIMEDe (ARCHivio Integrato di Microdati Economici e DEMo sociali), sviluppato a partire da marzo 2013², che si pone l'obiettivo di ampliare l'offerta informativa dell'Istat realizzando collezioni di dati elementari riferiti a famiglie e individui, da rendere disponibili all'utenza, utili alla ricerca sociale ed economica, alla programmazione territoriale e settoriale, alla valutazione delle politiche pubbliche a diversi livelli territoriali (Garofalo, 2014).

Attualmente, sono prodotte – attraverso l'integrazione di dati di fonte prevalentemente amministrativa- e diffuse con cadenza annuale, tre collezioni di microdati: “Popolazioni che usano un territorio e mobilità per lavoro/studio”, “Precarietà lavorativa” e “Condizioni socio-economiche delle famiglie”. “Popolazioni che usano un territorio e mobilità per lavoro/studio” ha come obiettivo la determinazione della popolazione insistente, costituita dall'insieme degli individui iscritti in anagrafe in Italia e degli individui non iscritti ma che lavorano o studiano in Italia³. Tra le componenti quantificate a livello di comune ci sono gli individui con mobilità in entrata, in uscita, dentro lo stesso comune, e quelli senza segnali di mobilità. “Precarietà lavorativa” ha lo scopo di analizzare l'universo dei lavoratori precari, descrivendo aspetti socio-anagrafici ed economico-lavorativi sia in un determinato periodo sia seguendone le trasformazioni nel tempo. “Condizioni socio-economiche delle famiglie” rappresenta una collezione di microdati relativa all'universo delle famiglie residenti che consente di analizzare congiuntamente vari aspetti delle loro caratteristiche socio-economiche e individuare eventuali aspetti di fragilità. In questo capitolo vengono descritte le principali caratteristiche delle basi dati “Precarietà lavorativa” e “Condizioni socio-economiche delle famiglie”, utilizzate per le elaborazioni presentate nella seconda parte del Volume.

1.2 La base dati “Precarietà lavorativa”

1.2.1 Le caratteristiche dell'attività

La precarietà lavorativa è un fenomeno complesso sia sul versante normativo,

¹ Il capitolo è frutto della collaborazione degli autori Sara Casacci e Dario Ercolani. In particolare, il paragrafo 1.2 è da attribuire a Dario Ercolani (ercolani@istat.it). Il paragrafo 1.3 e l'Appendice 1 sono da attribuire a Sara Casacci (casacci@istat.it). Si ringraziano Giuseppe Garofalo e Roberta Vivio coordinatori delle attività del progetto.

² Le attività sono attualmente realizzate nell'ambito dell'iniziativa “Costruzione e aggiornamento basi di dati per l'analisi territoriale” del MEB – Servizio Disegno dei processi e supporto al sistema dei registri dell'Istat.

³ Sono esclusi i non residenti non registrati negli archivi del Lavoro o Studio considerati: i non occupati, gli iscritti ad università straniere sul territorio italiano, quelli irregolarmente presenti sul territorio italiano.

sia su quello definitorio. Il lavoro precario è un concetto che riguarda aspetti oggettivi della condizione lavorativa di un individuo e 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, non esiste una definizione univoca di precarietà lavorativa riconosciuta a livello internazionale. 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.

1.2.2 Popolazione di riferimento e principali definizioni adottate

Il fenomeno è stato studiato a partire dall'analisi delle norme in tema di mercato del lavoro, nonché dall'esame di indagini e sistemi informativi implementati sia all'interno che all'esterno dell'Istat: il quadro normativo esaminato ha permesso di identificare le forme contrattuali di lavoro previste dall'ordinamento; le indagini statistiche analizzate hanno fornito conoscenze sulle informazioni raccolte riguardo le attività lavorative degli individui; i sistemi informativi esplorati hanno agevolato la ricognizione delle fonti informative disponibili e utilizzabili nel progetto.

La popolazione di riferimento è dunque costituita dagli individui che nel periodo di riferimento sono titolari di uno o più contratti di lavoro, anche svolti contemporaneamente. È stato considerato **lavoro precario**:

- il lavoro dipendente a termine;
- il lavoro prestato in base a rapporti di collaborazione per svolgere attività aventi caratteristiche di subordinazione;
- il lavoro autonomo consistente in attività aventi caratteristiche di subordinazione.

È quindi definito **lavoratore precario** l'individuo che, in un determinato periodo di tempo, svolge esclusivamente lavori definiti precari.

1.2.3 Le fonti di dati utilizzate

Le fonti di microdati sul lavoratore alla base dell'attività hanno natura fiscale e previdenziale e sono le seguenti.

Il principale contenuto informativo dell'integrazione è rappresentato dalla fonte denominata **Asia-Occupazione**⁴: si tratta di una base di microdati integrati per impresa e occupato, costruita ai fini dell'implementazione del Censimento Industria e Servizi 2011, nonché dell'aggiornamento del Registro Istat Asia-Imprese attive.⁵ La struttura⁶ 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 solida-

⁴ È costituita 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 dove l'attività lavorativa viene svolta).

⁵ 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.

⁶ Fonti integrate in Asia-Occupazione: 1. Lavoratori dipendenti (Inps – Emens/Dmag/Cig pagamento diretto/Ex Enpals Sport e Spettacolo, e Gestione Dipendenti Pubblici); 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).

rietà, nonché alcune variabili demografiche del lavoratore.

La fonte Inps dei **Lavoratori Autonomi in Agricoltura** riguarda gli imprenditori agricoli (coltivatori diretti e imprenditori agricoli professionali), ossia coloro che esercitano un'attività diretta alla coltivazione del fondo, alla silvicoltura, all'allevamento di animali e attività connesse. L'archivio contiene i dati contributivi relativi a tali soggetti.

L'archivio Inps dei **Lavoratori Domestici** contiene dati contributivi 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.

L'archivio Inps dei **Voucher di lavoro** (ex lavoro accessorio).

Le fonti descritte sono state analizzate, al fine di estrarre, elaborare e ottenere il seguente output: una base di dati corrispondenti ai singoli lavoratori, rappresentati dalla tipologia lavorativa principale svolta durante il periodo di riferimento allo studio. È stata usata una chiave multipla (variabili: Anno, Lavoratore, Tipologia lavorativa principale), dove, nel caso di molteplicità di attività di lavoro nel periodo di riferimento, la tipologia lavorativa principale è individuata in base al *ranking* che sintetizza le condizioni di precarietà lavorativa ricostruite (Prospetto 1.1).

Prospetto 1.1 – La classificazione della condizione di precarietà lavorativa

| Tipologia lavorativa | Altre caratteristiche osservate | Condizione di precarietà |
|---|---|--------------------------|
| 1. Dipendente a tempo indeterminato - Full Time | <ul style="list-style-type: none"> · Presenza di Cig · Mono/pluri committenza | Lavoro stabile |
| 2. Dipendente a tempo indeterminato - Part Time | <ul style="list-style-type: none"> · Presenza di Cig · Mono/pluri committenza | |
| 3. Autonomo | <ul style="list-style-type: none"> · Mono/pluri committenza | Lavoro precario |
| 4. Dipendente a tempo determinato | <ul style="list-style-type: none"> · Presenza di Cig · Mono/pluri committenza | |
| 5. Collaboratore | <ul style="list-style-type: none"> · Mono/pluri committenza | |
| 6. Lavoratore domestico | <ul style="list-style-type: none"> · Mono/pluri committenza | |
| 7. Voucher di lavoro | <ul style="list-style-type: none"> · Mono/pluri committenza | |

1.2.4 I principali contenuti informativi della base dati

La costruzione della collezione di dati 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 integrazione, omogeneizzazione e trasformazione.

Le principali caratteristiche lavorative di interesse ricostruite per il mese di ottobre sono: numero di datori di lavoro; numero di attività lavorative svolte; condizione contrattuale principale; tipologia contrattuale principale; condizione contrattuale principale anno precedente; intensità lavorativa mensile attività principale; intensità lavorativa mensile complessiva; presenza di segnale di Cassa Integrazione Guadagni e/o Contratto di solidarietà.

L'intensità lavorativa si configura come una misura di "presenza" annuale dell'individuo nelle fonti sull'occupazione analizzate. La quota annuale è computata a sua volta come media dell'intensità lavorativa mensile (Cfr. Capitolo 2). La presenza dell'occupato nelle fonti sul lavoro dipendente e parasubordinato (Inps incluse Gestione dipendenti pubblici, Gestione separata, lavoro domestico e *voucher*) è connessa al numero di settimane del mese per le quali si osserva un versamento di contributi previdenziali (a prescindere dall'orario giornaliero per i dipendenti); la presenza nelle fonti sul lavoro autonomo dipende dal periodo mensile di

iscrizione dell'occupato in Anagrafe Tributaria e nelle Camere di commercio o dal versamento di contributi nelle specifiche Casse previdenziali di lavoro autonomo (Artigiani e commercianti, Agricoltura). Osservate tutte le posizioni lavorative (fonti) che un soggetto può sostenere in un determinato periodo, l'intensità lavorativa mensile complessiva è calcolata sommando le intensità di tutte le posizioni (con somma massima unitaria); a livello annuale, l'intensità lavorativa è il rapporto tra la somma delle intensità mensili e il numero di mesi. L'output descritto è stato elaborato al fine di mettere in evidenza le potenzialità delle fonti integrate riguardo all'esplorazione statistica del fenomeno della "Precarietà lavorativa". Le elaborazioni hanno riguardato: analisi *cross section*; analisi dei flussi di entrata/uscita dalla condizione di occupato; analisi longitudinali e prime analisi sui redditi.

1.2.5 Le integrazioni con altre informazioni

Nel suo complesso, il progetto ArchIMEDe comprende più linee di attività in grado di esplorare diversi fenomeni mediante l'analisi dei contenuti informativi presenti in Istat. Tale prerogativa ha dato l'opportunità di creare delle sinergie: per ognuno dei prodotti è stato possibile sfruttare risultati di attività svolte nell'ambito degli altri. Nello specifico, alla base "Precarietà lavorativa" sono state integrate alcune informazioni reddituali derivanti da "Condizioni socio-economiche delle famiglie"⁷ al fine di effettuare analisi *ad hoc*.

Tra le classi di occupati in relazione alle quali è stata demarcata la precarietà lavorativa compaiono le collaborazioni, incluse quelle autonome con Partita Iva, che presentano caratteri 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 Partita 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.⁸ Tra gli sviluppi della base informativa realizzata sulla precarietà lavorativa, sono stati individuati i soggetti con Partita Iva potenzialmente mono-committenti, predisposti in un dataset strutturato al fine di analizzarne il grado di mono-committenza e ne sono state osservate le trasformazioni nel tempo. L'individuazione dell'universo delle Partite Iva (unità statistiche di riferimento) è stato effettuato a partire dal sistema informativo Asia-Occupazione. Per lo studio della mono-committenza sono state utilizzate diverse fonti amministrative acquisite da Istat. Il contenuto informativo più rilevante è presente in due fonti principali: (i.) il Modello 770 che i sostituti d'imposta (committenti) hanno l'obbligo di presentare per dichiarare i compensi erogati a persone e/o imprese con cui hanno avuto rapporti di lavoro e (ii.) il Mo-

⁷ Tali dati per l'anno 2011 provengono dalla Banca Dati Reddittuale del Ministero dell'Economia e delle Finanze.

⁸ 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. Tale 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 per cento 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.

dello Unico Persone Fisiche da cui si estraggono informazioni relative al reddito complessivo percepito dai lavoratori autonomi con Partita Iva.

1.3 La base dati “Condizioni socio-economiche delle famiglie”

1.3.1 Ricognizione delle esigenze informative e della letteratura

Una delle prime attività volte alla definizione dei contenuti informativi della base di dati è stata la ricognizione delle esigenze informative degli utenti. La specificazione del fabbisogno informativo è stata effettuata attraverso un'attività di confronto con alcune strutture dell'Istat e con alcuni soggetti Sistan. La rassegna della letteratura sul tema, unitamente all'analisi delle informazioni disponibili negli archivi amministrativi presenti in Istat, ha permesso di individuare quattro dimensioni ritenute significative, utili alla descrizione delle condizioni socio-economiche delle famiglie e alla eventuale identificazione di situazioni di vulnerabilità:

- *Reddito*: i redditi (insieme alla ricchezza) influenzano la possibilità delle famiglie di acquistare beni e servizi, determinando la loro resilienza da eventi avversi (perdite di lavoro, malattie a lungo termine, ecc.).
- *Lavoro*: questa dimensione rileva l'aspetto quantitativo della partecipazione al mercato del lavoro. Permette di evidenziare situazioni a maggior rischio di esclusione sociale, sottolineando gli effetti di una bassa intensità lavorativa.
- *Istruzione*: un alto livello di istruzione è positivamente correlato con standard di vita più elevati; inoltre aumenta la possibilità di trovare lavoro (OECD, 2010; Boarini et al., 2008; Sianesi e Van Reenen, 2003), e di adottare stili di vita sani (Miyamoto e Chevalier, 2010; de Looper e La Fortune, 2009).
- *Struttura della famiglia*: alcune strutture familiari hanno maggiori probabilità di sperimentare situazioni di povertà (McLanahan 1985; Cancian e Reed, 2002).

1.3.2 Le fonti utilizzate per la costruzione della base dati

Individuate le aree informative necessarie a coprire le dimensioni in esame sono state selezionate le fonti di natura amministrativa e censuaria già acquisite dall'Istat (o prodotti realizzati dall'Istat nell'ambito di altri processi).

Gli archivi amministrativi elementari (e le altre fonti) entrati nel processo ed i relativi contenuti informativi utilizzati (o derivati) sono riportati nel prospetto 1.

In generale, ciascuna fonte amministrativa acquisita è sottoposta ad alcune procedure di analisi volte alla sua validazione prima dell'utilizzo:

- analisi dei metadati, acquisizione di informazioni relative alla normativa e alla stabilità delle procedure che regolano la produzione della fonte amministrativa;
- verifica della presenza di valori anomali, fuori dominio o mancanti;
- controlli di stabilità temporale su numerosità degli oggetti dell'archivio e sulla distribuzione delle variabili di interesse.

1.3.3 Le fasi del processo di integrazione delle fonti di input

I dati di input del processo sono messi a disposizione dal Servizio RDD (Fonti amministrative e integrazione dei registri) dell'Istat attraverso la creazione di un

sistema informativo *ad hoc*, costituito da tabelle fisiche e viste logiche in cui sono memorizzate le informazioni provenienti dai diversi archivi amministrativi, derivato dal Sistema Integrato di Microdati (Sim) dell'Istituto. Il Sim contiene infatti un ricco patrimonio informativo, in termini di unità e di variabili; inoltre garantisce l'integrazione attraverso l'attribuzione di codici identificativi univoci validi nel tempo (Ambroselli, 2015). Non tutte le fonti di input risultano tuttavia integrate in SIM: esse vengono acquisite direttamente dalle strutture che le producono e integrate successivamente attraverso procedure di *linkage*. Il processo di costruzione della base dati⁹ prevede due livelli distinti, ma interconnessi, su cui vengono effettuate le procedure: individui e famiglie.

L'unità individuo si configura come l'unità di primo livello, dal momento che le informazioni sono raccolte ed elaborate in primo luogo su tale unità¹⁰. In seguito, attraverso apposite procedure, le informazioni disponibili a livello individuale sono riportate al livello familiare. Infine, le basi di dati vengono finalizzate e viene effettuata la selezione delle variabili per la diffusione¹¹.

1.3.4 Universo e unità di analisi

L'universo di riferimento è costituito dall'insieme delle famiglie i cui componenti risultano iscritti nelle Liste Anagrafiche dei Comuni (LAC) alla data di riferimento. L'unità di analisi è quindi la famiglia anagrafica, intesa come insieme degli individui residenti in famiglia (non sono considerati gli individui residenti in convivenza) e definita come l'insieme di persone legate da vincoli di matrimonio, parentela, affinità, adozione, tutela o affettivi, coabitanti ed aventi dimora abituale nello stesso comune¹². Nelle indagini campionarie sulle famiglie condotte dall'Istat l'unità di rilevazione è costituita, invece, dalla famiglia di fatto.

1.3.5 Le principali variabili socio-demografiche e relative all'occupazione

Le variabili provenienti dai singoli archivi amministrativi o dall'integrazione di più fonti sono create per derivazione da variabili amministrative tramite funzioni di trasformazione/derivazione di tipo deterministico (Wallgren e Wallgren, 2014).

Variabili demografiche e tipologia familiare

Le Lac contengono un *set* di informazioni riferite ai singoli componenti delle famiglie (data di nascita, sesso, stato civile, relazione di parentela rispetto all'intestatario del foglio di famiglia) per mezzo delle quali è possibile, da un lato, caratterizzare gli individui in funzione delle principali variabili demografiche (sesso, età, stato civile, cittadinanza italiana o straniera) e, dall'altro, articolare la composizione delle famiglie in termini di numerosità e tipologia. La tipologia familiare è stata determinata in funzione della combinazione di alcune modalità della relazione di parentela¹³ (Cfr. Capitolo 2). La qualità delle variabili demografiche e della tipologia e numerosità delle famiglie è strettamente dipendente dalla qualità delle informazioni delle liste anagrafiche da cui viene derivata e può essere inficiata

⁹ Il processo di costruzione della base dati "Condizioni socio-economiche delle famiglie" è stato ingegnerizzato dal servizio Progettazione e sviluppo dei sistemi per la produzione statistica (ITC) della Direzione Centrale per le Tecnologie Informatiche e della Comunicazione dell'Istat.

¹⁰ Nel caso in cui l'unità degli archivi amministrativi di input sia diversa dall'individuo (p. es. il trattamento pensionistico, la posizione lavorativa) si opera una riconduzione preliminare ad unità individuo.

¹¹ La diffusione dei dati effettuata agli utenti esterni all'Istituto prevede una selezione di variabili aventi come unità di riferimento le famiglie anagrafiche, relativamente ai soli comuni con più di 5.000 abitanti.

¹² Regolamento Anagrafico (legge 24 Dicembre 1954, n. 1228; DPR 30 Maggio 1989, n. 223).

¹³ Per maggiori dettagli sulla procedura di costruzione della tipologia familiare si veda Garofalo 2014, § 4.2.2.

da mancate regolarizzazioni anagrafiche di situazioni familiari mutate (Donati 1991) o da comportamenti “opportunistici” degli individui (ad esempio le residenze fittizie che non consentono di definire in modo corretto le unità familiari).

Livello di istruzione (Titolo di studio) e iscrizione scolastica ed universitaria

L’informazione relativa al titolo di studio dei componenti della famiglia deriva dal Censimento della Popolazione del 2011 e dai successivi aggiornamenti desunti dai dati Miur¹⁴. L’integrazione si rende necessaria in quanto le fonti Miur forniscono delle informazioni “di flusso” sui titoli di studio conseguiti ad una certa data, mentre il Censimento della Popolazione fornisce uno stock informativo rilevato al 9 ottobre 2011. Il titolo di studio è mancante per la popolazione non presente in percorsi di istruzione in Italia e non censita¹⁵. L’informazione relativa all’iscrizione scolastica o universitaria, per il singolo individuo, indica la sua presenza, rispettivamente, nell’Anagrafe Studenti¹⁶ e nell’Archivio degli iscritti e delle iscrizioni universitarie (Cfr. Capitolo 2).

Occupazione nel mese di ottobre e intensità lavorativa

Le informazioni relative alla occupazione nel mese di ottobre (e l’eventuale condizione di lavoro precario) e all’intensità lavorativa sono derivate dalla base “Precarietà lavorativa” (Cfr. § 1.2, Capitolo 2).

1.3.6 La costruzione del reddito

Il reddito complessivo lordo è inteso come somma di diverse componenti (p. es. redditi da lavoro, da pensione, da capitale, ecc.), indipendentemente dall’eventuale tassazione e dalla tipologia della tassazione stessa (ordinaria, separata). Al fine di ricostruirne l’ammontare, la soluzione adottata prevede il recupero di informazioni che consentono il maggior dettaglio possibile delle voci del reddito. La ricognizione delle fonti amministrative disponibili in Istat ha permesso di selezionare le basi di input utili all’implementazione di tale soluzione. La Banca Dati Reddittuale del Ministero dell’Economia e Finanze costituisce l’archivio *master* da cui sono ricavate le informazioni relative alla maggior parte delle voci di reddito. L’integrazione degli altri archivi consente, da un lato, di recuperare alcune voci di reddito che altrimenti risulterebbero sottostimate (p. es. le pensioni esenti, una stima della retribuzione dei lavoratori domestici, il reddito da lavoro autonomo dei contribuenti minimi, alcuni trasferimenti pubblici non soggetti a tassazione) e, dall’altro, di riclassificare alcuni importi (p. es. trasferimenti pubblici quali disoccupazione e mobilità sono scorporati dal reddito da lavoro dipendente e sommati al reddito da trasferimenti pubblici). Le macro-voci di reddito che costituiscono l’output della procedura di costruzione del reddito (e che sommate rappresentano il reddito complessivo) sono le seguenti:

- reddito da capitale finanziario;
- reddito da capitale reale;
- reddito da lavoro autonomo;
- reddito da lavoro dipendente;
- reddito da pensione;

¹⁴ In particolare, viene utilizzata la Base dati integrata su Istruzione e Titoli di studio – BIT – prodotta dal Servizio Fonti amministrative e integrazione dei registri della Direzione Centrale per la Raccolta Dati.

¹⁵ Non sono disponibili i dati relativi ai percorsi di istruzione Afam (Alta formazione Artistica, Musicale e Coreutica) sia di I, II livello e del vecchio ordinamento, i percorsi di Istruzione e Formazione Professionale – Ifp – gestiti dalle regioni e le specializzazioni tecniche superiori – Ifts.

¹⁶ Nella fonte non sono presenti i dati degli alunni delle scuole della regione autonoma della Valle D’Aosta. Fino all’anno 2013 non erano presenti anche i dati degli alunni delle scuole della provincia di Bolzano.

- reddito da trasferimento privato;
- reddito da trasferimento pubblico.

Occorre precisare che alle variabili riferite a ciascuna componente di reddito corrisponde una definizione di tipo “amministrativo” permeata su concetti di tipo fiscale e retributivo, associabile a specifiche sotto-popolazioni di contribuenti e/o percettori di un reddito certificato¹⁷.

I prospetti in Allegato 1 riportano, per ognuna delle macro-voci di reddito le relative componenti e la fonte da cui queste ultime sono derivate. È opportuno sottolineare che i microdati della base dati “Condizioni socio-economiche delle famiglie – ArchIMEDe” non sono confrontabili con quelli diffusi da Eu-Silc. In primo luogo, la definizione di reddito (e la relativa classificazione in macro-voci) adottata da Eu-Silc rappresenta un adattamento al contesto italiano di quella internazionale riportata nel manuale di Canberra (Unece, 2011), mentre i redditi presenti nella base dati presentano dei disallineamenti rispetto alla definizione ufficiale, risentendo quindi di una diversa quantificazione (Cfr. Capitolo 2). In secondo luogo, la base dati include soltanto voci di reddito al lordo della tassazione.

A livello familiare, il reddito lordo è ottenuto sommando le relative voci di reddito percepite da tutti i componenti della famiglia anagrafica. Tale reddito è quindi trasformato in reddito lordo equivalente utilizzando la scala di equivalenza Ocse¹⁸ che tiene conto sia del numero che dell’età dei componenti della famiglia.

Prospetto 1.2 – Archivi amministrativi e collezioni di dati di input e principali contenuti informativi utilizzati.

| Fonti dati | Ente titolare | Contenuti informativi |
|--|---|--|
| Lista Anagrafiche Comunali (Lac) | Comuni italiani | Caratteristiche anagrafiche dei componenti della famiglia; dimensione e tipologia familiare |
| Banca Dati Reddittuale (Bbr) | Ministero dell’Economia e delle Finanze (Mef) | Selezione di informazioni presenti nelle dichiarazioni dei redditi (modelli Unico Persone Fisiche, 730, 770) |
| Casellario dei pensionati (a) | Istituto Nazionale della Previdenza Sociale (Inps) | Importi delle prestazioni pensionistiche di varia natura |
| Rapporti di lavoro domestico | Inps | Stima della retribuzione percepita dai lavoratori domestici |
| Uniemens | Inps | Importi degli assegni al nucleo familiare per i dipendenti del settore privato |
| Modello Unico Persone Fisiche | Agenzia delle Entrate | Reddito derivante dall’esercizio di attività commerciali o di arti e professioni per determinate tipologie di contribuenti |
| Trattamenti non pensionistici | Inps | Erogazioni temporanee in denaro a individui/famiglie (es: disoccupazione; Assegni al nucleo familiare) |
| Banca Dati <i>Voucher</i> ¹⁹ | Inps | Importi percepiti per lavoro accessorio |
| Censimento Popolazione 2011 | Istat | Livello di Istruzione |
| Anagrafe degli studenti | Ministero dell’Istruzione, dell’Università e della Ricerca (Miur) | Livello di Istruzione, partecipazione a percorsi formativi scolastici |
| Archivio degli iscritti e delle iscrizioni universitarie | Miur | Livello di Istruzione, partecipazione a percorsi formativi universitari |
| “Precarietà lavorativa – ArchIMEDe” | Istat | Occupazione |

(a) Elaborazione dell’archivio fornita dal Servizio Sistema integrato salute, assistenza, previdenza e giustizia (Direzione Centrale per le Statistiche Sociali e il Censimento della Popolazione).

¹⁷ I dati fiscali per definizione non includono l’evasione fiscale che statisticamente coinvolge prevalentemente alcune categorie di percettori (su questo punto Cfr. Marino e Zizza, 2012).

¹⁸ La scala di equivalenza Ocse modificata utilizzata per calcolare il reddito familiare equivalente è pari alla somma di più coefficienti individuali (1 per il primo adulto, 0,5 per ogni altro adulto e 0,3 per ogni componente di età inferiore a 14 anni).

¹⁹ La banca dati *voucher* è utilizzata a partire dalla collezione dei dati relativi al 2014.

CAPITOLO II

METADATI PER CAPIRE I DATI¹

di Daniela Ferrazza

2.1 Metadati e qualità dei dati

L'obiettivo del capitolo è quello di chiarire in quali termini e per quali scopi è proceduto il lavoro sui metadati nell'ambito delle analisi dei dati territoriali esposte nella terza e quarta parte del volume. I casi di studio hanno utilizzato dati amministrativi resi disponibili dal progetto ArchIMEDe (Cfr. Capitolo 1) allo scopo di sperimentarne usabilità e resa a supporto delle politiche territoriali.

I metadati svolgono un ruolo cruciale nell'ambito della produzione della statistica ufficiale e fanno parte in maniera consolidata delle strategie degli Istituti di statistica europei allo scopo di migliorarne la produzione in termini non solo qualitativi. I metadati – la loro produzione e diffusione – sono legati strettamente alla qualità dei dati statistici e assumono declinazioni particolari quando si riferiscono a dati amministrativi che nascono, infatti, senza finalità statistiche seguendo criteri e processi diversi. Tale diversità pone specifiche sfide in tema di metadati poiché, oltre alla funzione di supporto alla comprensione emerge anche quella di facilitazione della traduzione e del trasferimento – anche “fisico” – da un ambito all'altro. Nondimeno, sempre più frequentemente anche il settore privato adotta approcci sistematici alla esplicitazione e all'uso dei metadati con particolare riferimento allo sfruttamento strategico dei dati in possesso delle aziende/organizzazioni attraverso strategie di Data Warehousing (Cfr. Capitolo 2) e Data Mining.

La definizione più semplice e breve di metadati li descrive come “*dati che spiegano altri dati*” (Fröschl 1997), in linea generale si tratta di informazioni che consentono di comprendere meglio il significato e il processo di produzione sottostante il dato per favorire una corretta interpretazione e la crescita di una conoscenza condivisa e intersoggettiva, affidabile e riproducibile.

Il concetto viene declinato in almeno due modalità, diverse tra loro, e legate ai diversi filoni scientifico-professionali che usano e producono metadati: *metadati strutturali* – vicini al linguaggio dei *computer scientists* – e *metadati di contenuto* (qui di riferimento), maggiormente interessanti per i professionisti dell'informazione.

I *metadati strutturali*, rispondono all'obiettivo di identificare, collocare e descrivere formalmente i dati statistici (p.es. etichette, posizioni di variabili in un tracciato, formati dei dati, liste dei codici, ecc.), mentre i *metadati di contenuto* li descrivono da un punto di vista semantico e consistono spesso in stringhe di testo libere. In questa sede ci si concentrerà maggiormente sui secondi poiché rispetto allo scopo di sviluppare dati e metodologie a supporto delle politiche territoriali, la disambiguazione e l'approfondimento semantico dei concetti/variabili diventano di cruciale importanza. I *metadati strutturali* saranno invece considerati nell'ambito

¹ Il capitolo riprende risultati del Gdl Metadati del progetto SPoT (WP “Griglia metadati progetto SPoT”). Sono stati membri del Gdl i colleghi Ballabio S., Carbonara M., Cariello S., Casacci S., Ferrazza D. (coord.), Succi R.

della definizione dell'architettura dei dati e della loro interrelazione (Cfr. Capitolo 4).

Il legame tra metadati e qualità delle statistiche ufficiali è stretto. Solitamente le principali dimensioni in cui il concetto di Qualità viene solitamente declinato (Dion, 2007; Eurostat, 2015) sono Rilevanza, Accuratezza, Tempestività, Accessibilità, Coerenza e Interpretabilità. Quest'ultima riguarda proprio la disponibilità di informazioni supplementari – metadati – per interpretare e utilizzare appropriatamente i dati. Tali informazioni si riferiscono a concetti, variabili, modalità e classificazioni, metodologia di raccolta.

Con riferimento all'uso di dati amministrativi anche la dimensione della coerenza di questi con altre informazioni statistiche richiede uno sforzo di esplicitazione di concetti e definizioni operative per cercare una concordanza sostanziale o, quantomeno, per chiarire le eventuali diversità nella consapevolezza che il concetto di coerenza non implica necessariamente una piena consistenza dal punto di vista numerico/quantitativo.

Il “perimetro” dei metadati è variabile, maggiori e/o diverse informazioni possono rendersi necessarie per la comprensione del dato a seconda della prospettiva da cui lo si legge e del fine che ci si propone. Le due macro categorie di soggetti interessati ai metadati – produttori e utilizzatori di statistiche – sono al loro interno ulteriormente articolate e includono ruoli e necessità diverse (i.e. statistici, metodologi, responsabili dei processi di produzione, esperti informatici, oppure *policy makers*, ricercatori, cittadini, ecc.). A questi diversi profili corrispondono diversi livelli di conoscenza pregressa, la varietà dei loro bisogni richiede non tanto la definizione di un set minimo di metadati, quanto una loro progettazione *ad hoc*.

Nel presente lavoro l'obiettivo dei metadati è esplicitare informazioni sostanziali e, in parte, di processo destinate a figure già alfabetizzate rispetto allo sfruttamento dei dati d'archivio, che necessitano di approfondimento sulle principali definizioni operative legate alle variabili e sulle conseguenze – di processo e di risultato – che esse implicano nell'analisi dei dati.

Nelle statistiche da indagine, le definizioni operative dei concetti vengono stabilite *ex ante* e considerate nella progettazione di tutte le fasi del processo di produzione statistica. Usualmente, per garantire la comparabilità dei dati nel tempo e nello spazio si consolida l'applicazione di alcune definizioni cardine che vengono utilizzate convenzionalmente a livello internazionale. Invece, per i dati amministrativi tipicamente le definizioni operative del produttore (ovvero dell'Ente titolare) non sempre (o meglio quasi mai) corrispondono a quelle in uso nella statistica ufficiale e, inoltre, sono soggette a mutamenti legati per esempio a cambiamenti normativi che possono essere più difficili da intercettare, perché manifestano i loro effetti negli archivi solo successivamente e richiedono un'attenzione particolare per la loro tracciatura. Anche solo per questa ragione, l'attenzione ai metadati per i dati amministrativi deve essere alta. Se poi si considera che i dati amministrativi sono ormai parte integrante di molti ambiti del sistema delle statistiche ufficiali, finora in gran parte costituite da indagini campionarie, emerge chiaramente la necessità di approfondire ulteriormente tali aspetti.

2.3 Gli ambiti dei metadati

I lavori presentati nella terza e quarta parte si basano su diverse annualità dei dati ArchIMEDe, dal 2012 al 2014, a seconda delle finalità delle analisi e dalle va-

lutazioni fatte dai ricercatori a tale proposito. Più precisamente, per ciascuna annualità sono disponibili sia i file di dati elementari riferiti agli individui sia il file di dati elementari riferito alle famiglie, derivato per aggregazione dai dati elementari (Cfr. Capitolo 1 e 4). Inoltre, è opportuno tenere distinte, idealmente, da un lato le variabili *originali* da fonte amministrativa e le variabili derivate in ArchIMEDe e, dall'altro, le variabili derivate nel progetto SPoT, di questo si troverà specifico riferimento nelle singole tabelle. Le variabili utilizzate da fonte ArchIMEDe, utilizzando fonti differenti, sono riconducibili a cinque ambiti semantici: Demografia, Territorio, Istruzione e formazione, Lavoro, Reddito².

2.3.1 Demografia

Le variabili demografiche sono utili per individuare i profili e le caratteristiche principali di individui (sesso, età, stato civile, ecc.) e famiglie (numero componenti, tipologie, ecc.) e sono state impiegate nelle analisi prevalentemente per osservare la diversa articolazione dei fenomeni oggetto di riflessione all'interno dei suddetti profili (Prospetto 2.1). In questo ambito ricadono principalmente variabili importanti o derivate dalle LAC (Liste Anagrafiche Comunali). L'universo di riferimento è costituito dall'insieme degli individui, e delle relative famiglie, iscritti nelle Liste Anagrafiche dei Comuni alla data di riferimento (01/01/2015 per la base dati del 2014, 01/01/2014 per la base dati del 2013 e 01/01/2013 per la base dati del 2012). L'unità di analisi privilegiata è quindi la famiglia anagrafica, intesa come insieme degli individui residenti in famiglia, e definita come l'insieme di persone legate da vincoli di matrimonio, parentela, affinità, adozione, tutela o affettivi, coabitanti ed aventi dimora abituale nello stesso comune. Una famiglia può essere costituita anche da una sola persona. La definizione di famiglia è quella contenuta nel Regolamento Anagrafico (legge 24 Dicembre 1954, n. 1228; DPR 30 Maggio 1989, n.223)³.

L'input del processo è costituito dalle LAC, dalle quali vengono selezionati gli individui ritenuti "*eleggibili*" (ovvero gli individui appartenenti a famiglie in cui nessun componente è sprovvisto di codici identificativi). Il file di dati sulle famiglie è derivato da quello individuale: i componenti di una famiglia sono identificati dalla combinazione delle variabili codice provincia, codice comune e codice famiglia. In termini di qualità dei dati, le LAC possono presentare criticità legate sia alla sovracopertura sia alla sottocopertura generate da: persone residenti in Italia ma che vivono all'estero e viceversa, mancate cancellazioni anagrafiche in particolare di stranieri (p.es. per emigrazione, comportamenti opportunistici dei cittadini che non applicano correttamente il principio della residenza anagrafica). Inoltre, ci possono essere problemi di sottocopertura dovute ai Comuni per cui le LAC non sono state acquisite dall'Istat. Rimandando ai singoli contributi di analisi per la gestione pragmatica di questo aspetto, in generale esistono tecniche che consentono di imputare un valore fittizio della variabile di interesse (p.es. cartografie SEBS, Cfr. Capitolo 6).

² Ai fini della ricostruzione qui presentata si fa riferimento a: Casacci S. a cura di, "Nota metodologica – Condizioni socio-economiche delle famiglie – ArchIMEDe" (documentazione interna); Ercolani D., a cura di, "Nota metodologica – Precarietà lavorativa – ArchIMEDe"; "La procedura per il calcolo del reddito nell'ambito del progetto ArchIMEDe – Precarietà Familiare (redditi 2012) – Nota ad uso interno"; "La procedura per il calcolo del reddito nell'ambito del progetto ArchIMEDe – Precarietà Familiare (redditi 2013) – Nota ad uso interno";

³ Nelle indagini campionarie sulle famiglie condotte dall'Istat l'unità di rilevazione è costituita, invece, dalla famiglia di fatto, intesa come un insieme di persone coabitanti e legate da vincoli affettivi, di matrimonio, parentela, affinità, adozione e tutela (Cfr. Capitolo 1).

Prospetto 2.1 - Variabili ambito demografico

| Variabile | Etichetta | Note |
|-------------------|---|---|
| SESSO | Sesso | |
| STACIV | Stato civile | |
| CITTAD | Cittadinanza | |
| ETA | Età | |
| REL_PAR | relazione di parentela | <u>Modalità</u> : 1. Intestatario; 2. Coniuge/convivente dell'intestatario; 3. Figlio/a dell'intestatario e/o del coniuge/convivente; 4. Genitore (o coniuge/convivente del genitore) dell'intestatario; 5. Suocero/a dell'intestatario; 6. Fratello/sorella dell'intestatario; 7. Cognato/a dell'intestatario; 8. Genero/nuora (coniuge/convivente del figlio/a) dell'intestatario; 9. Nipote dell'intestatario (sono inclusi in questa modalità sia figlio/a di figlio/a che figlio/a di fratello/sorella); 10. Nonno/a dell'intestatario; 11. Altro (legame con l'intestatario non altrove specificato); 12. Ignoto. |
| NCOMP | Numero componenti nucleo familiare | |
| N_COMP_CLETA1 | Numero componenti nucleo familiare tra 0 e 14 anni | |
| N_COMP_CLETA4 | Numero componenti nucleo familiare oltre i 65 anni | |
| NCOMP_EQUIVALENTI | Numero di componenti (adulti) equivalenti (Denominatore usato per il calcolo del reddito equivalente – vedi Prospetto 2.4.) | La variabile è costruita sulla base di una variabile di lavorazione NCOMP_013 (tra 0 e 13 anni) e viene utilizzato per tenere in considerazione effetti virtuosi di scala piuttosto che effetti viziosi legati alla composizione del nucleo familiare e, in particolare, al rapporto tra adulti e minori fino ai 13 anni. |
| TIPFAM2 | Tipologia familiare | <u>Modalità</u> : Unipersonali, Coppie senza figli, Coppie con figli, Monogenitore, Altro, Non classificabile. |
| CITTAD_FAM | Cittadinanza familiare | <u>Modalità</u> : Famiglia con tutti i componenti italiani, Famiglia con tutti i componenti stranieri o apolidi, Famiglia con componenti di cittadinanza italiana e straniera. |

2.3.2 Territorio

Il Codice Istat comunale rappresenta l'unità minima di riferimento rispetto alla gerarchia territoriale. Grazie alla presenza di questa dimensione, cioè della sua valorizzazione per ogni singolo record/caso dell'archivio, è possibile svolgere analisi ad un livello di granularità territoriale più basso rispetto a quanto avviene con i dati da indagine. Inoltre, è possibile procedere per aggregazioni seguendo criteri adatti al contesto di analisi prescelto e/o al fenomeno di riferimento (p.es. confini amministrativi, ASL, Sistemi Locali del Lavoro, ecc.). La fonte di riferimento della variabile è la Lista Anagrafica Comunale (Prospetto 2.2).

Prospetto 2.2 - Variabili ambito territoriale

| Variabile | Etichetta | Note |
|-----------|------------------------|--|
| COD_COM | Codice Istat Comune | Per il livello comunale da utilizzare congiuntamente. Per tale |
| COD_PRO | Codice Istat Provincia | motivo spesso si trovano unite (p.es. COD_PROCOM). |

2.3.3 Istruzione

Le fonti amministrative relative all'istruzione sono l'Anagrafe degli studenti e l'Archivio degli iscritti all'università e delle iscrizioni universitarie del MIUR – Ministero dell'Istruzione, dell'Università e della Ricerca (Prospetto 2.3). Per quanto riguarda il titolo di studio la fonte è il Censimento della Popolazione del 2011, aggiornato all'anno di riferimento attraverso informazioni del MIUR (Cfr. Capitolo 1). Non sono attualmente integrate le fonti informative su formazione professionale, AFAM – Alta formazione artistica, musicale e coreutica, dottorati di ricerca e tirocini. Questo determina una sottostima generalizzata con riferimento al comparto della formazione.

Prospetto 2.3 - Variabili ambito istruzione

| Variabile | Etichetta | Note |
|--------------------|--|--|
| TIPO_SCUOLA | Tipo di scuola | Modalità: 1. Primaria, 2. Secondaria di I grado; 3. Secondaria di II grado. La variabile è valorizzata se l'individuo è studente da fonte Anagrafe studenti MIUR. Può esservi un problema di sottocopertura legata al fenomeno degli studenti transfrontalieri. È opportuno tenere presente che la popolazione non include gli iscritti a formazione professionale Inoltre si segnalano i casi di alcune scuole pubbliche gestite da enti diversi dallo Stato, a scuole ospedaliere e a istituti carcerari. Si segnala mancanza di informazione per la regione Valle D'Aosta. |
| FLAG_UNIVERSITARIO | Flag di iscrizione universitaria | La variabile assume valore 1 se l'individuo risulta iscritto all'Università, diversamente 0. Può esservi un problema di sottocopertura poiché la popolazione non include gli iscritti a formazione professionale, AFAM, Dottorati di ricerca e tirocini. |
| TITSTU_RIC | Titolo di studio | Modalità: 1.Nessun titolo e non sa leggere o scrivere; 2. Nessun titolo ma sa leggere e scrivere; 3.Licenza di scuola elementare (o valutazione finale equivalente); 4.Licenza di scuola media; 5.Diploma di scuola secondaria superiore di durata 2-3; 6.Diploma di scuola secondaria superiore di durata 4-5; 7.Diploma di accademia di belle arti, danza, arte drammatica, isia, ecc. Conservatorio (vecchio ordinamento); 8.Diploma universitario (2-3 anni) del vecchio ordinamento (inlude le scuole dirette a fini speciali o parauniversitarie); 9.Diploma accademico di Alta Formazione Artistica, Musicale e Coreutica (A.F.A.M) di I livello; 10.Laurea triennale (di I livello) del nuovo ordinamento; 11.Diploma accademico di Alta Formazione Artistica, Musicale e Coreutica (A.F.A.M) di II livello; 12.Laurea (4-6 anni) del vecchio ordinamento, laurea specialistica o magistrale a ciclo unico del nuovo ordinamento, laurea biennale specialistica (di I livello) del nuovo ordinamento. Per quanto riguarda il titolo di studio è stata rilevata una sottocopertura. In particolare, le informazioni sui titoli di studio conseguiti all'estero dopo il 2011, sui diplomati AFAM e sui qualificati dei percorsi di istruzione e formazione gestiti dalle Regioni non sono disponibili. Inoltre, l'informazione è mancante per la parte della popolazione in LAC non censita. |
| BASSA_ISTRUZIONE* | % di persone con basso livello di istruzione | Percentuale di persone iscritte in anagrafe di 15 anni e più che: non hanno ancora conseguito la licenza media (fino ai 19 anni); hanno conseguito al più la licenza media (fino ai 59 anni); hanno conseguito al più la licenza elementare (dai 60 anni in poi). |

* Variabile derivata in SPoT.

2.3.4 Lavoro

La qualità e la quantità del lavoro svolto da un individuo è cruciale per lo studio di diversi fenomeni socio-economici di interesse pubblico (Prospetto 2.4). La situazione lavorativa (Cfr. Capitolo 7), la condizione di Neet (Cfr. Capitolo 11), la

disparità di genere nel mercato del lavoro (Cfr. Capitolo 10) sono solo alcuni esempi di temi affrontati nelle analisi qui presentate che hanno utilizzato variabili riferite a questo ambito. Le fonti principali sono Istat (ASIA – DB Occupazione), Inps (Gestione Dipendenti pubblici, Archivio Autonomi Agricoltura, Archivio lavoratori domestici). Il campo di osservazione considerato comprende gli occupati regolari, cioè lavoratori dipendenti e parasubordinati che versano contributi al Fisco Italiano e lavoratori autonomi iscritti in Anagrafe tributaria.

Esistono alcune differenze nelle modalità di rilevazione delle variabili relative a lavoratori dipendenti e/o parasubordinati da un lato e lavoratori indipendenti dall'altro che devono essere tenute debitamente in considerazione per la corretta interpretazione delle informazioni. Infatti, i segnali contributivi dei lavoratori dipendenti e parasubordinati hanno cadenza settimanale, mentre quelli dei lavoratori indipendenti hanno cadenza annuale, questo ha un impatto in tutte le variabili in cui avviene una "contabilizzazione" del tempo di lavoro degli occupati.

Come anticipato (Cfr. § 2.2) i lavori presentati nella Parte IV del volume utilizzano sia variabili provenienti da ArchIMEDe (originarie e derivate), sia variabili derivate nell'ambito del progetto SPoT. Poiché le variabili derivate sono di tipo dirette e indirette – cioè costruite a loro volta su variabili derivate – per semplicità di esposizione ci si limiterà qui a fornire metadati al livello in cui sono state utilizzate nelle analisi. Questo garantisce un'adeguata comprensione semantica del contenuto informativo delle variabili, senza appesantire eccessivamente il livello tecnico di dettaglio sulle loro singole componenti.

Le diverse variabili reperibili in ArchIMEDe relative all'ambito *lavoro* differiscono soprattutto per il riferimento temporale e la capacità di "misurare" l'intensità dell'attività lavorativa in maniera continua oppure discreta, attraverso la semplice indicazione della presenza/assenza un segnale di attività nel corso dell'anno⁴.

Prospetto 2.4 - Variabili ambito lavoro

| Variabile | Etichetta | Note |
|-----------|------------------------------|--|
| POS_MESE | Quota mensile lavorata | La variabile è una misura di presenza mensile delle diverse fonti sull'occupazione disponibili. Per i lavoratori dipendenti e parasubordinati la presenza dell'occupato nelle fonti è connessa al numero di settimane del mese per le quali si osserva un versamento dei contributi. Per i lavoratori autonomi la presenza dell'occupato nelle fonti sul lavoro dipende dal periodo mensile di iscrizione in Anagrafe Tributaria e nelle Camere di Commercio oppure dal versamento dei contributi nelle specifiche Casse previdenziali di lavoro autonomo. L'intensità lavorativa mensile è ottenuta sommando le intensità di tutte le posizioni che un soggetto può sostenere in un determinato periodo, stabilite a partire dall'osservazione di tutte le fonti disponibili (Cfr. Capitolo 1) |
| POS_ANNO | Intensità lavorativa annuale | <u>Valore:</u> Numero mesi lavorati nell'anno/12. La variabile è costruita a partire dalla quota mensile lavorata e rappresenta il rapporto tra il totale di mesi lavorati durante l'anno di riferimento e il totale di mesi potenzialmente disponibili per attività lavorative. Assume valori compresi tra 0 (totale assenza di segnali lavorativi nell'anno) e 1 (partecipazione continuativa nel mercato del lavoro nel corso dell'anno). |

⁴ Il lavoro oltre confine verrà affrontato nel Capitolo 3.

| | | |
|---------------------|--|---|
| LAV_ANNO_PREC | Flag di segnale di attività lavorativa nel corso dell'anno | La variabile è costruita a partire dall'intensità lavorativa annuale e assume valore 1 se l'intensità annuale è diversa da 0. |
| CONDIZ_OTT_ANNOCORR | Condizione lavorativa principale ad ottobre | <u>Modalità:</u> 1. Dipendente a tempo indeterminato; 3. Autonomo; 4. Altro dipendente; 5. Dipendente a tempo determinato; 6. Collaboratore; 7. Lavoratore domestico; 8. Voucher ⁵ . La variabile è prodotta mediante l'uso di fonti diverse: per il lavoro dipendente e parasubordinato è stato possibile osservare i segnali contributivi mensili, mentre sul lavoro autonomo i segnali hanno cadenza annuale, pertanto la numerosità di questa tipologia di lavoratori a ottobre risulta sovrastimata. Tra le eventuali attività lavorative svolte si è individuata quella relativa alla tipologia contrattuale considerata più stabile. |
| FLAG_OCC | Flag occupato nel mese di ottobre | La variabile assume valore 1 se l'individuo risulta occupato nel mese di ottobre, altrimenti 0. |
| FLAG_LAVDIP* | Flag lavoratore dipendente | Il flag assume valore 1 se gli individui soddisfano, nell'anno di riferimento, la condizione di aver lavorato oltre una certa soglia e se contestualmente presentano una condizione lavorativa di lavoro dipendente o assimilabile (Dipendente a tempo determinato, lavoratore domestico, voucher). |

* Variabile derivata in SPoT.

2.3.5 Reddito

Le variabili relative al reddito sono presentate nel prospetto 2.5 e nel dettaglio descritte nelle note inserite. Il reddito individuale totale costituisce il fulcro logico di molte analisi condotte ed è dato dalla somma di reddito da lavoro, da pensione, da capitale, da trasferimenti pubblici e privati.

Prospetto 2.5 - Variabili ambito reddito

| Variabile | Etichetta | Note |
|------------------|--|--|
| REDD_LAVDIP3 | Reddito da lavoro dipendente ricalcolato | La variabile è calcolata come somma dei redditi da lavoro dipendente e assimilato, al lordo degli straordinari. Oltre ai lavoratori dipendenti in senso stretto sono inclusi anche i soci di cooperative di produzione e lavoro, i lavoratori socialmente utili, i co.co.pro., i co.co.co, nonché una stima della retribuzione dei lavoratori domestici. Sono inoltre compresi i premi di produzione assoggettati ad imposta sostitutiva. Sono scorporati alcuni importi che costituiscono un reddito da trasferimento pubblico, derivabili dall'archivio Inps dei Trattamenti Monetari non pensionistici (Cfr. Capitolo 1). È ricompresa nel calcolo una stima della retribuzione totale percepita dal lavoratore domestico, ottenuta a partire dalle informazioni sui versamenti Inps. |
| REDD_CALCULO4 | Reddito individuale totale | Reddito calcolato come somma di varie tipologie di reddito (da lavoro, da pensione, da capitale, da trasferimenti pubblici e privati) al lordo delle imposte. La variabile è derivata in ArchIMEDe da integrazione di diverse fonti Banca Dati Reddituale (BDR), Casellario dei pensionati, Rapporti di lavoro domestico, Uniemens, Modello Unico Persone Fisiche, Trattamenti non pensionistici (Cfr. Appendice 1). |
| REDD_FAM | Reddito familiare | Somma del reddito individuale dei componenti la famiglia anagrafica. |
| REDD_EQUIVALENTE | Reddito familiare equivalente (scala OCSE) | La variabile è costruita – secondo le indicazioni OCSE – come rapporto tra il Reddito familiare e il numero di componenti (adulti) equivalenti, pertanto si rapporta il reddito familiare a un fattore di scala usato per rendere equivalenti |

⁵ La modalità "Voucher" è codificata unicamente nel rilascio 2015, anno utilizzato nel Capitolo 9.

| | | |
|-------------------|----------------------|--|
| | | i redditi di famiglie di diversa ampiezza e composizione, in modo da tener conto dei diversi bisogni di minori e adulti e delle economie di scala che si realizzano con la coabitazione di più componenti. |
| FLAG_POVERTA_ITA* | Flag famiglia povera | Famiglie con reddito familiare equivalente inferiore al 60 per cento del reddito familiare equivalente mediano. |

* Variabile derivata in SPoT.

2.3.6 Ambiti misti

A partire dalle variabili afferenti ai diversi ambiti illustrati è possibile procedere per costruzione verso variabili/indici più complessi che sintetizzano al loro interno informazioni provenienti da aree differenti (Prospetto 2.6). Diverse variabili di questo tipo sono state utilizzate nelle analisi illustrate nel presente volume, sia – come anticipato sopra – derivate nell’ambito del progetto SPoT.

Prospetto 2.6 - Variabili ambiti misti

| Variabile | Etichetta | Note |
|------------------------------------|---|--|
| FLAG_NEET_F e FLAG_NEET_M | Persones di età compresa tra i 15 e 29 anni che non studiano e non lavorano | La variabile è definita per costruzione dall’incrocio di variabili relative all’ambito lavorativo (LAV_ANNO_PREC) e variabile relative all’istruzione. Gli individui che hanno un segnale di presenza lavorativa nel corso dell’anno di riferimento vengono considerate occupate; le persone che risultano iscritte a una scuola di ogni ordine e grado sono considerate studenti. |
| FLAG_LAVDIP_POV* | Flag lavoratore dipendente povero | Lavoratore dipendente con reddito da lavoro dipendente inferiore al 60 per cento del reddito mediano da lavoro dipendente. |
| CLASS_WP* | Classificazione working poor | Variabile sintetica che identifica una classificazione basata sulla combinazione tra situazione reddituale del singolo e situazione reddituale della famiglia in cui è inserito. Modalità: 00. Non working poor; 01. Lavoratore povero in famiglia non povera; 10. Lavoratore non povero in famiglia povera; 11. Lavoratore povero in famiglia povera. |
| BASSA_INTENSITA_LAVORATIVA* | % di persone residenti in famiglie a bassa intensità lavorativa | Percentuale di persone che risiedono in famiglie anagrafiche la cui intensità di lavoro è inferiore a 0,20, ossia incidenza di persone che vivono in famiglie dove le persone in età lavorativa (tra i 18 e i 59 anni, con l’esclusione degli studenti 18-24 anni) hanno lavorato per meno del 20 per cento del loro potenziale (con esclusione delle famiglie composte solo da minori, da studenti di età inferiore a 25 anni e da persone di 60 anni o più). |
| BASSA_DISPONIBILITA_REDDITI_LORDI* | % di persone residenti in famiglie a basso reddito | Percentuale di persone residenti in famiglie anagrafiche con reddito lordo equivalente pro-capite familiare minore del 60 per cento del reddito equivalente mediano nazionale sul totale delle persone residenti in famiglie anagrafiche. |

* Variabile derivata in SPoT.

CAPITOLO III

QUALITÀ DEI DATI AMMINISTRATIVI

FOCUS SUL MOVIMENTO FRONTALIERO NON RILEVATO

di Maria Elena Comune e Lorena Carmen Maria Viviano¹

3.1 Premessa

Lo studio del sistema formativo e del mercato del lavoro a partire da dati di fonte amministrativa ci pone di fronte a nuove sfide relative alla qualità del dato. Sono emerse, infatti, importanti distorsioni da studi effettuati in precedenza, soprattutto in alcune sottopopolazioni concentrate in specifici territori. Ad esempio, le stime dei fenomeni relativi agli ESL (Early School Leavers - Ballabio et al. 2016a, Ballabio et al. 2018b) e ai Neet² (Ballabio et al. 2016c, Ferrazza et al. 2017, Ballabio et al. 2018a) presentano valori elevati sui confini nazionali con la Svizzera, non del tutto coerenti con le aspettative basate sulla conoscenza del territorio. Tra le fonti di distorsione figurano i flussi frontalieri³, l'IeFP (Istruzione e Formazione Professionale) e i *missing data*⁴. In particolare, la componente dei frontalieri è la fonte di distorsione più consistente in alcune aree lombarde come è emerso nello studio dei Neet.

Per costruzione in ArchIMEDe figurano i dati delle dichiarazioni dei redditi trasmesse all'Agenzia delle Entrate da parte dei contribuenti, eccetto quello degli italiani che lavorano in Svizzera che sono esentati dal presentarle in Italia come espresso negli Accordi Italia – Svizzera del 3 ottobre 1974 (Cfr. Appendice 2). Per questo motivo ArchIMEDe non rileva gli spostamenti (quotidiani o settimanali) di coloro che pur risiedendo nei comuni di confine italo-svizzero (entro i 20 km)⁵, lavorano o studiano oltre confine. Questa mancanza determina una sovrastima della componente di coloro che non studiano e non lavorano (Neet) non corrispondente alla realtà di fatto. In questo capitolo l'attenzione viene rivolta al fenomeno dei frontalieri, mentre l'approfondimento sui *missing data* (Cfr. Appendice 3) e lo studio dell'IeFP (Cfr. Capitolo 8) sono in parte trattati in altre sezioni del volume.

L'obiettivo di questo capitolo è di provare a fornire una valutazione (qualitativa e quantitativa) sull'assenza informativa dei flussi frontalieri a livello locale. Il fenomeno dei frontalieri è difficilmente catturabile sia dalle statistiche ufficiali di tipo campionario (p.es. Rilevazione Continua Forze di Lavoro – RCFL – Istat) per

¹ Hanno collaborato alla stesura dell'Appendice 2 Simona Ballabio e Arianna Carra.

² Neet: Giovani di età compresa tra i 15 e i 29 anni che non lavorano e non studiano, cioè che non sono occupati disoccupati e inattivi e non sono inseriti in un percorso di educazione e/o formazione.

³ Frontalieri: Qualsiasi persona che si reca abitualmente (almeno una volta alla settimana) nel territorio di un paese limitrofo, diverso da quello in cui risiede, per esercitarvi un'attività lavorativa subordinata o autonoma.

⁴ A causa dell'assenza dell'informazione comunale o perché inizialmente errate le trasmissioni comunali, sono possibili sia *missing* parziali sia *missing* totali.

⁵ Per i residenti in Italia oltre i 20 km dal confine troverà applicazione l'articolo 15 della Convenzione contro le doppie imposizioni stipulata tra Italia e Svizzera, in base alla quale il reddito di lavoro dipendente prodotto in Svizzera dal lavoratore frontaliere viene tassato in Italia, beneficiando della franchigia di 7.500 euro per i redditi di lavoro dipendente prodotti all'estero in zone di frontiera.

problemi di rappresentatività (Ballabio et al. 2016b) sia – come anticipato – da ArchIMEDe a causa della attuale assenza di fonti secondarie adeguate. L'unica fonte in grado di quantificare il flusso suddiviso per provincia di provenienza è il dato del pendolarismo disponibile in occasione del Censimento della popolazione, con la limitazione rappresentata dalla cadenza decennale della rilevazione.

Rispetto alle applicazioni, la componente frontiera viene pertanto qui valutata indirettamente attraverso l'impiego dei dati censuari 2011. Non essendo da molti punti di vista ArchIMEDe omogeneo al Censimento della popolazione (p.es. per definizione operativa, metodo di raccolta, ecc.) si è ipotizzato di inserire nel Censimento le distorsioni note in ArchIMEDe relative ai frontalieri.

3.2 Fonti sui frontalieri

Stime attendibili e robuste sui flussi frontalieri (da e per la Svizzera) sono fornite trimestralmente dall'Ust (Ufficio Federale di Statistica svizzero) e annualmente dall'Ustat (Ufficio di Statistica del Canton Ticino; Comune et al. 2017)⁶.

Come si può osservare nella tabella 3.1, il dato sui frontalieri che si spostano per motivi di lavoro / studio dal Piemonte e dalla Lombardia (regioni di confine) verso la Svizzera di fonte censuaria (Censimento della popolazione del 2011) è molto prossimo alla stima fornita dall'Ust. È da sottolineare che il dato censuario permette di distinguere la provenienza territoriale dei frontalieri verso la Svizzera, consentendo di quantificare i flussi provinciali. D'altra parte, i dati censuari non permettono di distinguere i flussi frontalieri per cantone di destinazione, poichè il dato rilevato si riferisce al totale della Svizzera.

Il flusso dall'Italia verso il Ticino, cantone verso cui sono diretti circa il 90 per cento dei movimenti provenienti dall'Italia verso la Svizzera, è, invece, stimato da Ustat – Ufficio di Statistica del Canton Ticino. Dalla lettura dei dati (Tabella 3.1) emerge chiaramente la concentrazione del fenomeno in poche province (Como e Varese).

3.3 Frontalieri di 15-29 anni

Per meglio capire il celato fenomeno dei frontalieri il *benchmark* censuario risulta di aiuto per indagare sulle province insubriche, dal momento che i dati del pendolarismo consentono di calcolare esaustivamente quanti individui si recano per motivi di studio e di lavoro oltre confine, con frequenza giornaliera o settimanale.

In particolare, il dato censuario dei giovani di età 15-29 anni residenti in Lombardia è stato ripartito per provincia e per condizione professionale, secondo le modalità di risposta previste nel questionario censuario, tralasciando i possibili e specifici approfondimenti definatori sulle categorie presenti ("Casalinghe, Disoccu-

⁶ I dati sui frontalieri sono disponibili ai seguenti link:

- Ust: <https://www.bfs.admin.ch/bfs/en/home/statistics/work-income/employment-working-hours/employed-persons/swiss-foreign-nationals/cross-border-commuters.html>

- Ustat: <https://www3.ti.ch/DFE/DR/USTAT/index.php?fuseaction=temi.dati&p1=35&p2=151&p3=160&prold=159>

I dati sono resi disponibili dal Responsabile del settore economia dell'Ufficio di Statistica del Canton Ticino-Giubiasco (O. Gonzalez, email agli autori, 26 settembre 2016). Si tratta di stime calcolate utilizzando la Statistica dei frontalieri (Staf) dell'Ust e i rilasci del permesso di lavoro secondo il Sistema d'informazione centrale sulla migrazione (Simic) dell'Ufficio federale della migrazione di Berna.

pati, In cerca di prima occupazione, percettori di pensione e in altra condizione”). Anche nella fascia di età 15-29 anni il fenomeno dei frontalieri appare più consistente in poche province: l’82 per cento dei frontalieri della regione risiede a Varese e a Como (Tabella 3.3). Tuttavia osservando le incidenze sulla popolazione, si può notare come il fenomeno sia significativo in tre province: la percentuale di frontalieri risulta assente (pari a zero) nella maggior parte delle province lombarde, mentre Como Varese e Sondrio registrano flussi di frontalieri rispettivamente del 5 per cento, 4 per cento e 3 per cento (Tabella 3.3). Al di fuori della regione Lombardia, tuttavia andrebbe considerata anche la provincia di Verbano-Cusio-Ossola, unica provincia piemontese di confine i cui flussi frontalieri verso la Svizzera sono piuttosto consistenti (6 per cento).

Tabella 3.1 - Flusso di frontalieri per movimento (origine e destinazione) e fonte. 2011-2012

| Destinazione Origine | Svizzera | | | | Ticino | |
|-------------------------|--------------------------------------|---|--|--|--|--|
| | RCFL 2011 (Istat) ¹ | Censimento popolazione 2011 (Istat) ² | Statistiche dei frontalieri 2011 (UST) ³ | Statistiche dei frontalieri 2012 (UST) ³ | Statistiche dei frontalieri 2011 (USTAT) ⁴ | Statistiche dei frontalieri 2012 (USTAT) ⁴ |
| Italia | 37.000 | (a) 55.306 | 56.892 | 60.976 | 51.618 | 55.300 |
| Lombardia | 31.500 | 49.408 | n.d. | n.d. | 45.676 | 49.027 |
| Varese | 6.929 | 22.074 | n.d. | n.d. | 22.337 | 23.853 |
| Como | 18.879 | 20.709 | n.d. | n.d. | 20.797 | 22.288 |
| Sondrio | 1.515 | 3.993 | n.d. | n.d. | 326 | 349 |
| Milano | 2.288 | 1.201 | n.d. | n.d. | 1.627 | 1.934 |
| Lecco | 527 | 293 | n.d. | n.d. | 281 | 291 |

Fonte: Ballabio et al. (2016b) e Gonzalez (Cfr. Nota 6).

Note: (a) Il dato è riferito a Piemonte e Lombardia; 1. Stime campionarie annuali; 2. Dato censuario riferito al 9 ottobre 2011; 3. Stime annuali; 4. Stime con diffusione annuale. I dati provinciali sono riportati unicamente per i flussi frontalieri più consistenti. Pertanto la loro somma non coincide con il totale regionale.

Tabella 3.2 - Popolazione di 15-29 anni, per condizione professionale e provincia. Lombardia, 2011

| Provincia | Neet | | | | | | Studenti | Occupati | Totale |
|------------------|---------------|---------------|---------------------------------------|------------------------|------------------------|----------------|----------------|----------------|------------------|
| | Casalinghe | Disoccupati | In cerca di prima occu- pazione | Percettore pensione | In altra condizione | Totale | | | |
| Varese | 4.423 | 6.283 | 5.297 | 158 | 5.442 | 21.603 | 46.833 | 56.521 | 124.957 |
| Como | 2.768 | 3.585 | 3.167 | 88 | 3.409 | 13.017 | 31.488 | 40.107 | 84.612 |
| Sondrio | 857 | 1.140 | 556 | 48 | 1.119 | 3.720 | 9.645 | 14.080 | 27.445 |
| Milano | 11.793 | 20.196 | 16.275 | 1.066 | 18.339 | 67.669 | 163.376 | 181.054 | 412.099 |
| Bergamo | 7.232 | 6.944 | 6.135 | 282 | 5.989 | 26.582 | 57.258 | 82.776 | 166.616 |
| Brescia | 9.521 | 8.146 | 6.744 | 244 | 7.471 | 32.126 | 65.181 | 89.740 | 187.047 |
| Pavia | 3.078 | 3.462 | 3.238 | 171 | 3.770 | 13.719 | 26.670 | 30.881 | 71.270 |
| Cremona | 2.260 | 2.410 | 2.105 | 110 | 2.057 | 8.942 | 18.431 | 23.544 | 50.917 |
| Mantova | 3.188 | 2.480 | 2.146 | 81 | 2.606 | 10.501 | 19.323 | 26.601 | 56.425 |
| Lecco | 1.537 | 2.067 | 1.696 | 42 | 1.811 | 7.153 | 18.956 | 23.219 | 49.328 |
| Lodi | 1.423 | 1.499 | 1.410 | 31 | 1.318 | 5.681 | 11.648 | 15.472 | 32.801 |
| Monza e Brianza | 3.527 | 5.510 | 4.797 | 183 | 4.502 | 18.519 | 46.466 | 53.712 | 118.697 |
| Lombardia | 51.607 | 63.722 | 53.566 | 2.504 | 57.833 | 229.232 | 515.275 | 637.707 | 1.382.214 |

Fonte: Istat – Censimento della popolazione 2011.

Note: Definizioni operative del censimento della popolazione 2011: In occasione del Censimento del 2011, per la prima volta, alcune informazioni di carattere socio-economico sono state rilevate su base campionaria attraverso l’uso

di due tipi di questionario: uno in forma ridotta, con pochi quesiti, tra cui quelli indispensabili per la produzione dei dati richiesti dall'Unione Europea ad un elevato dettaglio territoriale, e uno in forma completa in cui sono state aggiunte le altre variabili previste nel piano di rilevazione. In particolare, nei comuni con una numerosità demografica fino a 20.000 abitanti tutti i cittadini hanno risposto alla forma completa del questionario. Nei comuni con una popolazione di più di 20.000 abitanti 1/3 dei cittadini presenti nella Lista anagrafica comunale (Lac) al 1 gennaio 2011 ha risposto alla forma completa del questionario, 2/3 alla forma ridotta. In tutti i casi, i cittadini non presenti nella Lac comunale al 1° gennaio del 2011 hanno risposto alla forma completa del questionario.

L'età è desunta dalla data di nascita rilevata nella lista A e nella sezione I (quesito 1.3) presenti sia nella forma ridotta sia nella forma completa del questionario. Al fine dello studio del fenomeno dei Neet sono stati considerati i soggetti di età compresa tra i 15 e i 29 anni.

I quesiti del questionario censuario (sezione 6) che rilevano la condizione occupazionale sono presenti sia nella versione completa sia in quella ridotta e permettono di definire le diverse modalità utili allo studio del fenomeno.

Le modalità considerate sono: **1. Occupati:** coloro che nella settimana precedente alla data di riferimento del Censimento (dal 2 all'8 ottobre 2011) hanno svolto almeno un'ora di lavoro oppure, se non lo hanno svolto, avevano comunque un lavoro dal quale erano assenti; **2. Non occupati: In cerca di occupazione:** tutti i non occupati, come da definizione precedente, che dall'11 settembre all'8 ottobre hanno cercato attivamente un lavoro alle dipendenze o hanno predisposto i mezzi per avviare un'attività in proprio e che, qualora si presentasse l'opportunità, sarebbero disposti ad iniziare un lavoro entro le due settimane successive. Rientrano in questa modalità sia i disoccupati, vale a dire coloro che hanno svolto un'attività lavorativa in passato, sia gli individui in cerca di prima occupazione, coloro cioè che non mai svolto un'attività lavorativa. **Inattivi non studenti:** tutti i non occupati e non disoccupati, come da descrizioni precedenti, che si definiscono percettori di una o più pensioni per effetto di attività lavorativa passata o percettore di reddito da capitale (per rendita da investimenti immobiliari o mobiliari) oppure casalingo oppure in altra condizione. **3. Studenti: tutti coloro che si definiscono studenti.** Sono stati classificati Neet i soggetti che non sono né occupati né studenti. La popolazione così circoscritta può essere confrontata con la quota di soggetti che al quesito 7.1, presente in entrambe le forme del questionario, dichiarano di non recarsi giornalmente al luogo abituale di studio o di lavoro perché non lavorano, non studiano e non frequentano corsi di formazione professionale. Nella sola versione completa del questionario, la condizione di studente può essere confrontata con le risposte fornite ai quesiti 5.11 e 5.12. Qui i soggetti dichiarano un'eventuale iscrizione a un corso regolare di studi oppure un'eventuale frequenza, dal 2 all'8 ottobre, di un corso di formazione o aggiornamento professionale. La popolazione dei frontalieri è stata invece definita da tutti i soggetti che dichiarano che il loro luogo abituale di studio e di lavoro, in cui si recano giornalmente, è all'estero (quesito 7.1, presente nella versione completa).

Tabella 3.3 - Popolazione frontiera di 15-29 anni, per provincia, 2011

| Provincia | Totale | % sul totale frontalieri | % sul totale popolazione |
|----------------------|---------------|--------------------------|--------------------------|
| Varese | 4.599 | 42% | 4% |
| Como | 4.406 | 40% | 5% |
| Sondrio | 923 | 8% | 3% |
| Milano | 421 | 4% | 0% |
| Bergamo | 127 | 1% | 0% |
| Brescia | 86 | 1% | 0% |
| Pavia | 38 | 0% | 0% |
| Cremona | 18 | 0% | 0% |
| Mantova | 9 | 0% | 0% |
| Lecco | 122 | 1% | 0% |
| Lodi | 14 | 0% | 0% |
| Monza e Brianza | 159 | 1% | 0% |
| Lombardia | 10.922 | 100% | 1% |
| Verbano-Cusio-Ossola | 1.144 | / | 6% |

Fonte: Istat, Censimento della popolazione 2011.

3.4 Risultati: la rilevanza dei flussi frontalieri per le stime dei Neet

A partire dai dati censuari, sono state calcolate l'incidenza dei Neet e l'incidenza dei Neet "aumentata" dai frontalieri sulla popolazione totale 15-29 anni complessiva per le province lombarde rispetto alla popolazione presente sia per il Censimento della popolazione sia in ArchIMEDe. Relativamente ai Neet presenti in ArchIMEDe vengono fornite tre misure: i Nieet (*Not Income, in Employment*,

Education or Training)⁷, i Neet riferiti all'intero anno 2012 e i Neet riferiti al mese di ottobre 2012 (in analogia al mese di riferimento censuario). I dati, riferiti agli ultimi anni disponibili, sono stati utilizzati ipotizzando che il differente riferimento temporale (2011 per il Censimento e 2012 per ArchIMEDe) non rappresenti una criticità, trattandosi di valutazioni sulla struttura della popolazione che tende a rimanere stabile in anni contigui. Le incidenze ottenute sono state ordinate secondo il rango (dal valore più elevato al più basso) per individuare come cambia la posizione di una provincia al modificarsi dell'indicatore di riferimento, nel caso in cui si riesca a cogliere il movimento dei frontalieri (dati censuari) e nel caso in cui tale informazione manchi (ArchIMEDe). Dai risultati (Tabella 3.4), emerge che l'inclusione dei frontalieri (i.e. Neet+frontalieri) modifica significativamente la distribuzione delle province portando nelle prime posizioni Varese e Como, in cui, insieme alla presenza dei Neet, è nota la rilevante componente frontiera (Comune e Viviano, 2016). Il confronto tra il *rank* censuario e quello di ArchIMEDe mostra anche altri cambiamenti di posizione che andrebbero ulteriormente indagati. I più rilevanti sono emersi nelle province di Milano, Cremona, Monza e Sondrio.

Infine, è stato usato lo strumento delle *Box plot map* (Cfr. Capitolo 6) per l'individuazione di aree critiche, ponderando sia rispetto ai singoli comuni sia rispetto all'intera popolazione comunale, con un differente impatto sui piccoli comuni. Nelle prime due cartografie si rappresentano i Neet calcolati sui dati censuari. Si può osservare che le aree con maggiore concentrazione dell'indice sono presenti nel pavese (Figure 3.1a e 3.1b). Se ai Neet si aggiungono i frontalieri (Figure 3.2a e 3.2b), si osserva una maggiore incidenza del fenomeno nell'area di confine con la Svizzera (province di Varese, Como e Sondrio). Le cartografie successive (Figure 3.3a e 3.3b), invece, rappresentano l'incidenza dei Nieet quantificati a partire dall'archivio ArchIMEDe. In questo caso l'area critica si restringe e si concentra maggiormente nell'area delle province di Como e Varese.

Tabella 3.4 Incidenza e rank dei Neet, per provincia. Lombardia, 2011 e 2012

| Provincia | 2011 | 2011 | 2012 | 2012 | 2012 | 2011 | 2011 | 2012 | 2012 | 2012 |
|------------------|--------------|--------------------|--------------|--------------|--------------|-----------------|--------------------|-----------------|-----------------|-----------------|
| | Cens. Pop. % | Cens. Pop. % | Archl-MEDe % | Archl-MEDe % | Archl-MEDe % | Cens. Pop. Rank | Cens. Pop. Rank | Archl-MEDe Rank | Archl-MEDe Rank | Archl-MEDe Rank |
| | Neet | Neet + Frontalieri | Nieet | Neet anno | Neet ottobre | Neet | Neet + Frontalieri | Nieet | Neet anno | Neet ottobre |
| Sondrio | 13,55 | 16,92 | 12,42 | 18,17 | 26,46 | 13 | 9 | 13 | 13 | 11 |
| Lecco | 14,50 | 14,75 | 15,02 | 20,44 | 26,21 | 12 | 13 | 12 | 11 | 13 |
| Bergamo | 15,95 | 16,03 | 15,86 | 20,26 | 26,40 | 9 | 11 | 11 | 12 | 12 |
| Monza e Brianza | 15,60 | 15,74 | 16,38 | 21,84 | 27,79 | 10 | 12 | 10 | 7 | 9 |
| Cremona | 17,56 | 17,60 | 16,47 | 21,09 | 27,71 | 3 | 5 | 9 | 10 | 10 |
| Lodi | 17,32 | 17,36 | 17,05 | 21,51 | 28,21 | 4 | 7 | 8 | 9 | 8 |
| Brescia | 17,18 | 17,22 | 17,09 | 21,54 | 28,31 | 6 | 8 | 7 | 8 | 7 |
| Mantova | 18,61 | 18,63 | 17,51 | 22,29 | 29,81 | 2 | 4 | 6 | 6 | 5 |
| Lombardia | 16,58 | 17,37 | 17,52 | 22,70 | 29,39 | 7 | 6 | 5 | 5 | 6 |
| Milano | 16,42 | 16,52 | 17,81 | 23,46 | 30,45 | 8 | 10 | 4 | 4 | 4 |
| Pavia | 19,25 | 19,30 | 19,31 | 24,37 | 31,73 | 1 | 3 | 3 | 3 | 3 |
| Varese | 17,29 | 20,97 | 20,03 | 25,58 | 32,00 | 5 | 1 | 2 | 2 | 2 |
| Como | 15,38 | 20,59 | 20,30 | 25,96 | 32,40 | 11 | 2 | 1 | 1 | 1 |

Fonte: Istat, Censimento della popolazione 2011 e ArchIMEDe 2012.

⁷ Ballabio et al. (2018a).

Figura 3.1a - Neet 2011 (peso unitario)

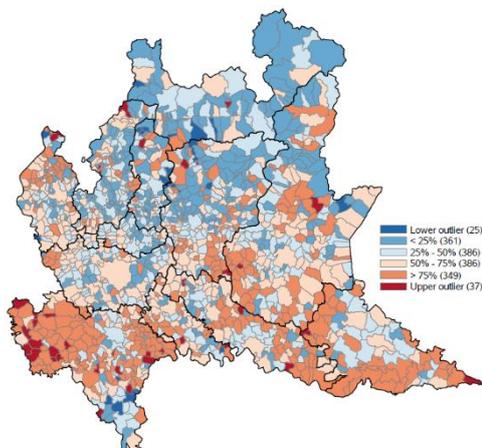


Figura 3.1b - Neet 2011 (peso popolazione target comunale)

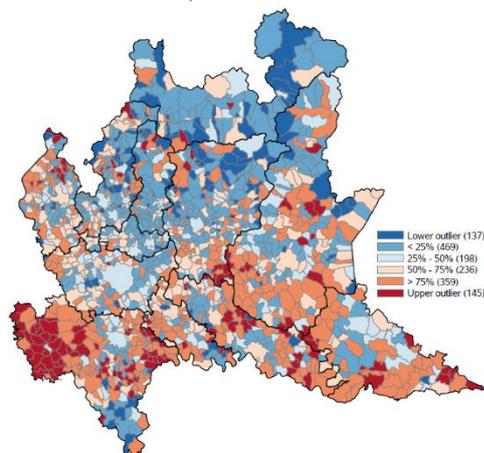


Figura 3.2a - Neet e frontalieri 2011 (peso unitario)

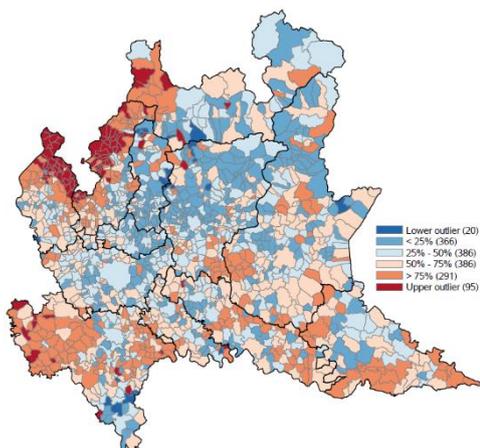


Figura 3.2b - Neet e frontalieri 2011 (peso popolazione target comunale)

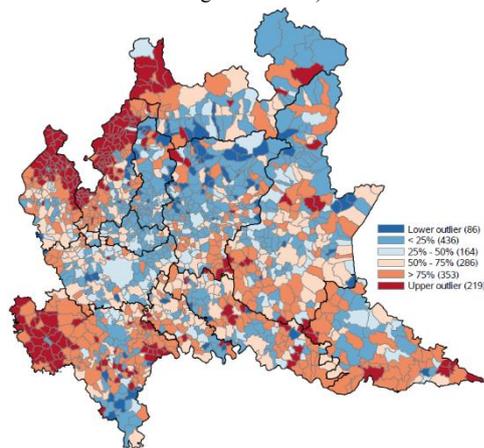


Figura 3.3a - Nieet 2012 (peso unitario)

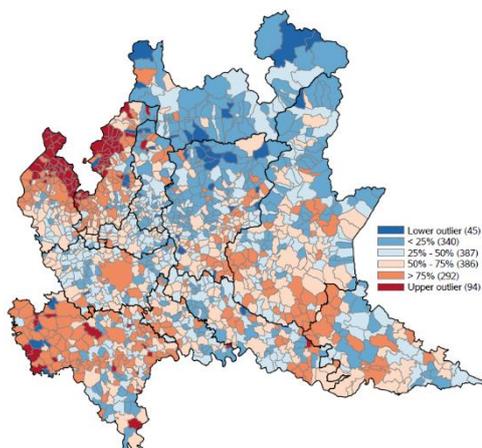
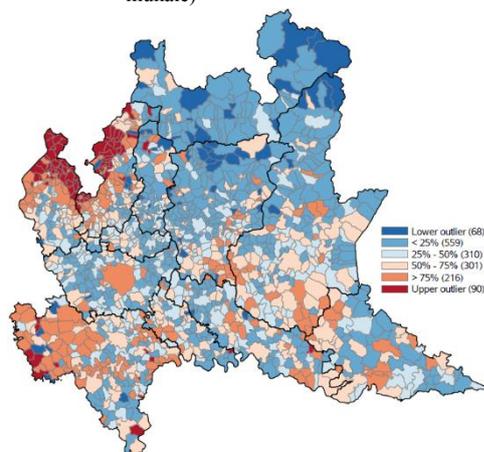


Figura 3.3b - Nieet 2012 (peso popolazione target comunale)



Fonte: Istat, Censimento della popolazione 2011 e ArchIMEDe 2012.

3.4.1 Analisi dei ranghi

A scopo descrittivo vengono rappresentate le differenze di rango delle province nel confronto dei dati censuari rispetto ai dati di ArchIMEDe riferiti ai Neet e ai Neet+frontalieri (Tabella 3.5). Successivamente si passa al calcolo della correlazione fra ranghi:

$$\rho = 1 - \frac{6 - \sum_i d_i^2}{n(n^2 - 1)} \quad (1)$$

n rappresenta il numero delle province, mentre d sono le differenze di ranghi.

Attraverso il calcolo della correlazione fra ranghi (Tabella 3.6), si può osservare come la presenza dei frontalieri sommata ai Neet del Censimento determina una struttura più simile a quella di ArchIMEDe nel caso in cui si considerano i Neet quantificati nel mese di ottobre (0,78). Ciò indica che il flusso dei frontalieri ha una sua rilevanza nella conoscenza del fenomeno nelle province insubriche. Gli sfasamenti di posizione per alcune province (p.es., Milano e Monza) andrebbero indagati con approfondimenti successivi (p.es., introducendo i dati relativi all'Istruzione e Formazione Professionale).

Si osserva infine che la considerazione dei *missing* non ha fornito indicazioni particolarmente significative in quanto non sono cambiate le strutture territoriali (Cfr. Appendice 2).

Tabella 3.5 - Differenza rank Neet Censimento della popolazione vs ArchIMEDe, per provincia. Lombardia, 2011 e 2012

| Provincia | Neet (Cens. Pop.) | Neet (Cens. Pop.) | Neet (Cens. Pop.) | Neet + fronta- lieri (Cens. Pop.) | Neet + fronta- lieri (Cens. Pop.) | Neet + fronta- lieri (Cens. Pop.) |
|-----------------|-------------------------|-------------------------|----------------------|---|---|---|
| | vs | vs | vs | vs | vs | vs |
| | Nieet | Neet anno | Neet ottobre | Nieet | Neet anno | Neet ottobre |
| Sondrio | 0 | 0 | 2 | 4 | 4 | 2 |
| Lecco | 0 | 1 | 1 | 1 | 2 | 0 |
| Bergamo | 2 | 3 | 3 | 0 | 1 | 1 |
| Monza e Brianza | 0 | 3 | 1 | 2 | 5 | 3 |
| Cremona | 6 | 7 | 7 | 4 | 5 | 5 |
| Lodi | 4 | 5 | 4 | 1 | 2 | 1 |
| Brescia | 1 | 2 | 1 | 1 | 0 | 1 |
| Mantova | 4 | 4 | 3 | 2 | 2 | 1 |
| Milano | 4 | 4 | 4 | 6 | 6 | 6 |
| Pavia | 2 | 2 | 2 | 0 | 0 | 0 |
| Varese | 3 | 3 | 3 | 1 | 1 | 1 |
| Como | 10 | 10 | 10 | 1 | 1 | 1 |

Fonte: Istat, Censimento della popolazione 2011 e ArchIMEDe 2012.

Tabella 3.6 - Correlazione fra ranghi, per provincia. Lombardia, 2011 e 2012

| Neet (Cens Pop) vs Nieet | Neet (Cens Pop) vs Neet anno | Neet (Cens Pop) vs Neet ottobre | Neet + frontalieri (Cens Pop) vs Nieet | Neet + frontalieri (Cens Pop) vs Neet anno | Neet + frontalieri (Cens Pop) vs Neet ottobre |
|--------------------------------------|--|--|--|--|---|
| 0,43 | 0,32 | 0,40 | 0,77 | 0,68 | 0,78 |

Fonte: Istat, Censimento della popolazione 2011 e ArchIMEDe 2012.

3.5 Conclusioni

Il fenomeno dei frontalieri determina una distorsione significativa nel trattamento dei dati di archivio. Gli approfondimenti territoriali vanno considerati con cautela lì dove riguardano le province di confine dove il fenomeno è più consistente (Cfr. Appendice 2). Una possibile soluzione è quella di integrare le fonti mancanti con le informazioni acquisite dagli uffici di statistica elvetici (Ust e Ustat) per i Cantoni in cui i movimenti sono maggiori (i.e. Ticino, Grigioni e Vallese) ipotizzando un loro coinvolgimento per la fornitura e la costruzione di archivi.

Attualmente l'Istat collabora a un progetto che coinvolge diversi Paesi europei per la quantificazione dei flussi a partire dalle informazioni presenti in varie fonti amministrative (i.e. *Border regions data collection project*⁸).

⁸ Agreement n. 2016.CE.16.BAT.105 European Commission – Directorate General for Regional and Urban Policy (Coordinamento per l'Italia: L. Franconi).

INTRODUZIONE ALLA PARTE SECONDA

METODI

di Paola Maddalena Chiodini

Grazie alle fonti amministrative, si hanno oggi grandi quantità di dati a disposizione della Statistica ufficiale per l'analisi dei fenomeni socio-economici. Questa nuova prospettiva determina la necessità dell'applicazione e dell'evoluzione non solo delle tecniche statistiche ma anche di quelle più specificatamente legate alle tecnologie dell'informazione. Tali basi dati implicano la necessità di gestione di dati di carattere multidimensionale che devono essere classificate ed organizzate per la loro fruibilità ed utilizzo.

Con il crescente impiego dell'informatica e lo sviluppo di applicativi dedicati, le operazioni di raccolta e archiviazione dei dati sono diventate più efficienti consentendo la gestione ed il trattamento di moli di dati un tempo impensabili. Al contempo le attività gestionali si sono anche molto articolate e diversificate. Si è passati, infatti, dall'uso esclusivo dei tradizionali sistemi di gestione di basi di dati (dbms), prioritariamente finalizzati all'archiviazione dei dati, all'impiego di applicazioni specializzate funzionali all'analisi dei dati giungendo alla realizzazione di veri e propri *data warehouse*. Questi ultimi consistono in raccolte di dati *integrate, orientate al soggetto, variabili nel tempo e non volatili* (Inmon, 1996). La logica e la modalità di archiviazione dei dati devono essere funzionali al processo di *decision making*, in tal senso il *data warehouse* si deve adattare ed essere funzionale all'utilizzatore che deve accedere ed usare le informazioni in esso contenute. Il contributo di Ballabio, Ferrazza e Verrecchia sui metodi per la gestione dei dati da fonte secondaria argomenta l'importanza del *data warehousing end-to-end* per l'acquisizione, la trasformazione e l'organizzazione dei dati finalizzati allo studio, all'analisi e alla rappresentazione dei fenomeni a supporto delle politiche locali.

Se da un lato la capacità di organizzare e ottimizzare grandi quantità di dati rappresenta il primo passo da compiere per lo studio di fenomeni complessi, dall'altro lato sono necessari adeguati metodi per l'estrazione e la sintesi delle informazioni. In alcuni casi quali la misurazione dei fenomeni multidimensionali relativi alla povertà, la vulnerabilità, ecc. trattati nel volume, può risultare particolarmente complessa per via della natura stessa del dato e delle leggi distributive che lo definiscono. Le tecniche statistiche consentono la costruzione di indici di sintesi, tuttavia queste semplificazioni sono di per sé soggettive e comportano, inevitabilmente, il rispetto di requisiti in conflitto tra loro, oltre a basarsi su considerazioni di natura empirica. Il trade-off nell'ambito dell'aggregazione dei dati risiede nella capacità di sintesi degli indicatori e al contempo nella minimizzazione della perdita d'informazione. Il contributo di Mazziotta sui metodi per la costruzione di misure a supporto delle politiche territoriali, inerente agli indici sintetici, illustra alcuni metodi utilizzati nel volume per l'aggregazione di variabili che operativizzano fenomeni multidimensionali (p.es. MPI, AMPI).

A conclusione del ciclo di raccolta, gestione, analisi e sintesi delle informazioni nell'ambito di un processo decisionale, un ulteriore aspetto riguarda la diffusione

e rappresentazione efficace delle informazioni e degli indici necessari allo studio dei fenomeni. In questo contesto, per dati di natura territoriale, il cartogramma è uno strumento particolarmente efficace per l'illustrazione dei fenomeni locali. In particolare, la possibilità di rappresentare indicatori territoriali attraverso cartografie "temperature map" – che presentano la distribuzione dei fenomeni con intensità di colore differente – permette al lettore o al decisore di aver informazioni immediatamente fruibili. Per lo studio della distribuzione territoriale dell'incidenza di un fenomeno sono individuabili diversi strumenti cartografici che rispondono a differenti domande conoscitive. In particolare, sono ritenute adatte per rappresentare l'intensità e la modalità di concentrazione o dispersione di un fenomeno a livello territoriale le *Quartile map*, le *Box plot map*, le *Local indicator spatial association (LISA) cluster map* (Anselin, 1994) e le *Spatial empirical bayesian smoothed rate (SEBS) box plot map* (Anselin, 2004) presentate nel capitolo VI, sui metodi per la rappresentazione di dati territoriali di supporto per le politiche locali. Queste tecniche sono particolarmente interessanti in quanto costituiscono modalità di rappresentazione dei dati di fonte amministrativa che, pur mantenendo un elevato contenuto informativo, garantiscono informazioni coerenti con le statistiche ufficiali.

Questa parte del volume dedica quindi attenzione alla presentazione di soluzioni metodologiche inerenti l'approccio gestionale e informativo finalizzato alla fruibilità delle informazioni provenienti da fonte secondaria, i metodi per la costruzione di indicatori ed infine i metodi grafici rivolti alla diffusione di informazioni territoriali di supporto alle politiche dei governi locali. La rosa di metodologie presentate sono altresì di supporto per lo sviluppo dei casi di studio illustrati nel volume.

CAPITOLO IV

GESTIRE I DATI AMMINISTRATIVI

di Simona Ballabio, Daniela Ferrazza e Flavio Verrecchia

4.1 Introduzione

L'Istituto Nazionale di Statistica (Istat) ha adottato il *software* Oracle per l'archiviazione delle basi dati di interesse dei progetti inseriti nel Programma Statistico Nazionale (PSN). Oracle ha come obiettivo principale l'archiviazione e l'accessibilità del dato, che certamente sono aspetti fondamentali in considerazione dell'attuale fase di transizione della statistica ufficiale verso l'orizzonte degli archivi amministrativi. Questo software consente un'ampia gamma di funzionalità per il trattamento dei dati, ciononostante l'Istat, non solo per tradizione, utilizza anche altri applicativi statistici per le analisi e lo studio dei fenomeni (p.es. SAS).

In questa sede, gli applicativi utilizzati con lo scopo di “*pre-trattamento*” e successiva analisi e rappresentazione del dato sono:

- l'applicativo statistico SAS che è stato utilizzato per l'implementazione di un processo di *data warehousing* cosiddetto “*end-to-end*”. Infatti, è stato impiegato per l'accesso ai dati Oracle e, anche grazie all'ampia scelta di strategie di analisi che rende disponibili, per le elaborazioni e le stime ed infine per la rappresentazione delle informazioni (p.es. attraverso cartografie);
- il software GeoDA che è stato utilizzato per alcune applicazioni di statistica spaziale. GeoDA, inizialmente sviluppato dal *team* di Luc Anselin del Laboratorio di Analisi Spaziale dell'Università dell'Illinois di Urbana-Champaign, è un *software freeware* e *open source*, utile per l'analisi dei dati spaziali e per la geovisualizzazione. Il programma fornisce un'interfaccia grafica di facile utilizzo per i metodi di analisi esplorativa dei dati spaziali, come le statistiche di autocorrelazione spaziale per i dati aggregati e l'analisi di regressione spaziale di base per i dati puntuali e poligonali.

4.2 Il processo di *data warehousing end-to-end*

La filosofia di riferimento, di interesse negli ambiti di *secondary data collection*, ha la sua centralità nel *data warehouse* (DW) quale strumento capace di trattare e sistematizzare dati, originariamente raccolti secondo prospettive e finalità differenti, per realizzare sistemi informativi orientati alla qualità e alla conservazione della conoscenza e del controllo dei processi produttivi del dato stesso.

I dati prodotti quindi sono (e diventano), in primis, strumenti di controllo e solo successivamente sono, eventualmente, considerabili per la diffusione di statistiche ufficiali.

L'approccio *end-to-end* si caratterizza per la considerazione complessiva del processo di *delivery* delle informazioni dal *back-end* – fase di accesso alle fonti – al trattamento, alla sintesi del dato e alla diffusione delle informazioni (i.e. *front-end*). Secondo questo approccio il DW diventerebbe il luogo da cui possono nascere le statistiche ufficiali da fonte amministrativa. Chiaramente, in questa prospettiva, alle

attività di controllo del dato dovrebbero seguire quelle di correzione e di retroazione per rivedere e completare le informazioni selezionate da caricare nei *data mart* – sottoinsiemi di dati di un DW appartenenti a specifici e ben determinati ambiti di analisi – rivolti alla diffusione. Tali *Feedback* sono necessari anche per il passaggio dall'ambiente di sperimentazione alla produzione. Come è necessaria l'implementazione di accurati e dettagliati metadati per l'utilizzo del dato (Cfr. capitolo 2).

4.2.1 Il Data warehouse

Centrale rispetto al processo *end-to-end* è il *data warehouse* che può essere definito come una collezione di dati integrata, orientata a soggetti di interesse, rappresentativa della dinamica temporale e non volatile, a supporto del processo decisionale (Inmon, 1996). Più in particolare:

- per integrata si intende l'integrazione dei dati raccolti o acquisiti da fonti esterne. Integrazione che riguarda anche l'uso di metodi di codifica uniformi e di trattamenti finalizzati all'omogeneità semantica delle variabili e delle relative modalità, vale a dire alla omogeneità dei metadati strutturali (Cfr. capitolo 2). Il DW restituisce una visione unificata e consistente di tutte le basi di dati che sono fonte del processo. In questa sede, facendo il *back-end* riferimento alla *secondary data collection* da fonte amministrativa, questo aspetto risulta di particolare interesse;
- per orientata a soggetti di interesse si intende l'orientamento del DW a popolazioni target specifiche. In un DW, infatti, i dati sono archiviati in modo da essere facilmente letti o elaborati, tuttavia, l'obiettivo non è quello di minimizzare la ridondanza mediante la normalizzazione e l'integrazione dei dati funzionale alle applicazioni aziendali ma, al contrario, quello di denormalizzare al fine di fornire dati organizzati per soggetto (p.es. i comuni, i residenti, i lavoratori, le donne, i giovani, ...) e quindi per favorire la produzione di informazioni. Dalla progettazione per funzioni si arriva dunque ad una modellazione multidimensionale dei dati (Cfr. § 4.2.2);
- per rappresentativa della dinamica temporale si intende che nel DW sono contenute informazioni relative ad una determinata popolazione target e riferite ad un specifico intervallo temporale. I dati contenuti in un DW sono pertanto aggiornati fino ad una certa data. Ciò differisce da quanto si verifica in un sistema operativo¹, nel quale i dati sono aggiornati, ma non consoni a fornire un quadro storico del fenomeno analizzato;
- per non volatile si fa riferimento sia all'assenza di sofisticate procedure di inserimento e cancellazione del dato prevedendo l'accesso in sola lettura per l'utenza sia alla non temporaneità delle informazioni contenute.

Nel caso del progetto ArchIMEDe (Cfr. Capitolo 1) – prima – e nel *datawarehouse* SPoT (Cfr. § 4.3) – poi – il *repository* è stato costruito integrando diverse fonti, con un *focus* su individui e famiglie residenti, e considerando diversi rilasci temporali di basi dati con periodicità annuale.

¹ Anche detti sistemi transazionali sono sistemi gestionali adatti a processare transazioni ovvero sequenze di operazioni elementari. Spesso in questo ambito si fa direttamente riferimento a sigle quali OLTP (*On-Line Transaction Processing*) e ERP (*Enterprise Resource Planning* – pianificazione delle risorse d'impresa), sistemi di gestione che integrano tutti i processi di business rilevanti per una organizzazione.

4.2.2 Dal modello E-R allo star schema

Il modello *entity-relationship* (E-R), che consente la rappresentazione concettuale dei dati ad un livello di astrazione elevato (Chen, 1976), può essere utilizzato per la rappresentazione delle entità e delle relazioni dei sistemi transazionali o di archiviazione delle informazioni di tipo secondario. Lo *star schema* (i.e. schema a stella), invece, è la tipologia più semplice di schema per la rappresentazione della struttura dati di un DW. La rappresentazione multidimensionale è anche di utilità per la progettazione e implementazione dei DW. La struttura di base consiste in una tabella di fatti a cui si aggiungono le relative tabelle delle dimensioni.

4.3 Progetto SPoT-DeBEST

Nell'ambito del progetto, pur non utilizzando interfacce specifiche per la realizzazione e gestione di DW, è stato adottato l'approccio e il *frame* metodologico di riferimento per l'emulazione delle fasi tipiche del processo di *data warehousing end-to-end*.

4.3.1 La progettazione concettuale

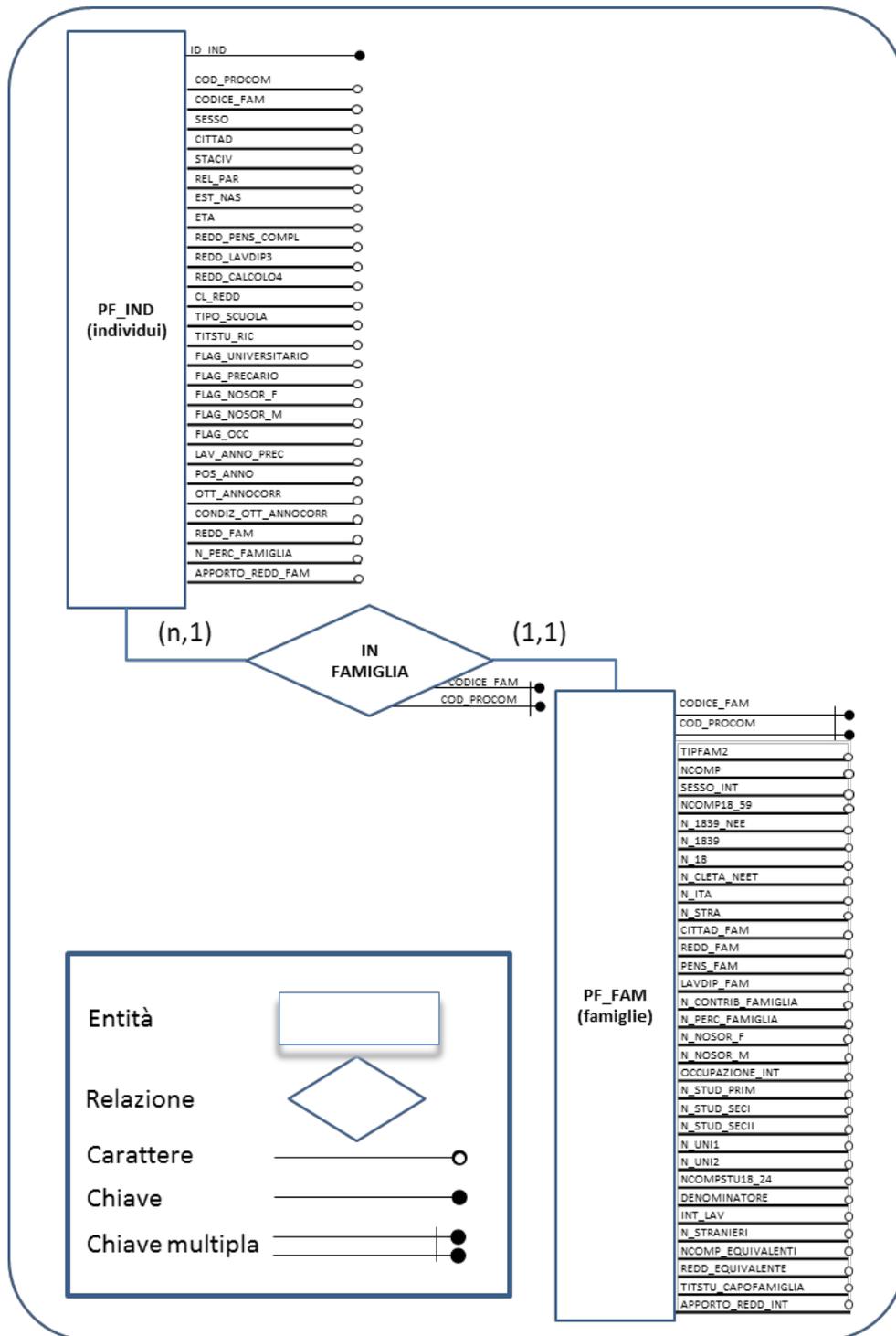
Il modello concettuale utilizzato nel progetto SPoT dell'Istat è condizionato in quanto molta parte dell'informazione è già integrata nelle basi dati fornite dal progetto ArchIMeDe. In ogni caso, un diagramma semplificato relativo a individui e famiglie è disponibile in figura 4.1, rimandando invece – per approfondimenti relativi alle strutture di rappresentazioni – a volumi che specificatamente trattano l'argomento (Istat, 1989). In particolare, grazie alla chiave multipla composta dal codice famiglia (CODICE_FAM) e dal codice del comune di residenza (COD_PROCOM) di fonte anagrafica è possibile usare il legame relazionale al fine di utilizzare congiuntamente dati sugli individui e sulle relative famiglie (p.es. Reddito individuale vs Reddito equivalente pro-capite familiare).

Lo schema multidimensionale (Figura 4.2) di fatto rappresenta una base dati parzialmente aggregativa nel senso che per produrre le statistiche risulta necessario utilizzare la dimensione temporale come filtro, distinguendo i funzionali opportuni per individui e famiglie.

Scelto il processo da modellare (p.es. struttura e dinamica della popolazione residente del Nord-Ovest), grazie allo schema a stella si possono rappresentare e distinguere: i *fatti*, eventi riferibili alla popolazione target; le relative *dimensioni* (p.es. stato civile, condizione lavorativa, ecc.) con modalità discrete che qualificano i fatti; le *gerarchie* (p.es. territorio amministrativo), che sono alberi i cui nodi sono dimensioni e i cui archi modellano associazioni uno-a-molti tra coppie di dimensioni. Le *measure* (p.es. reddito, numero di componenti, ecc.) invece descrivono aspetti quantitativi di un fatto².

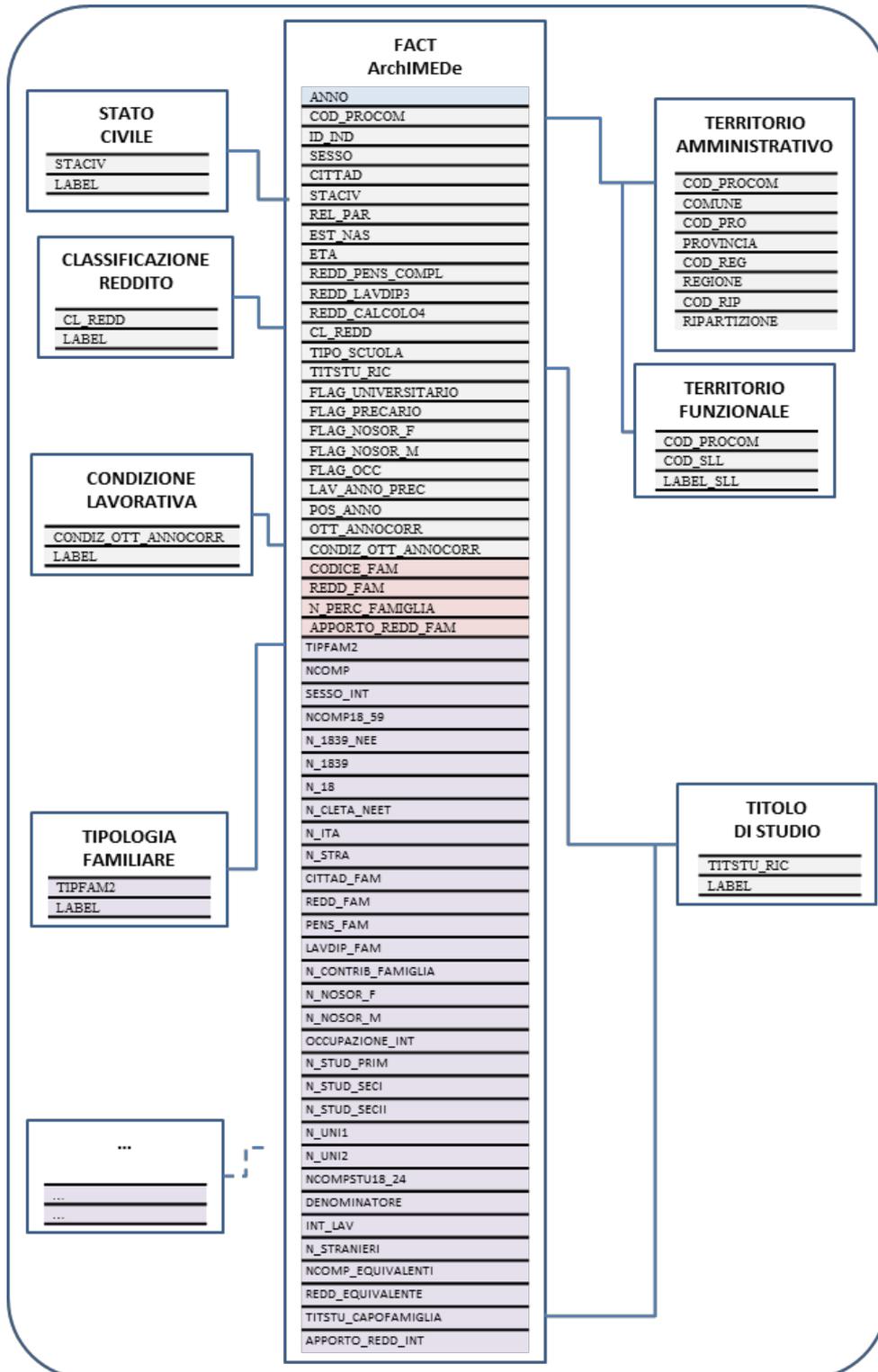
² Per una descrizione articolata si rimanda al Capitolo 2, § 2.3.

Figura 4.1 - Diagramma E-R semplificato del progetto SPoT-DeBEST



Note: **INDIVIDUI:** ID_IND | Codice individuo; COD_PROCOM | Codice Istat Provincia Comune di residenza; CODICE_FAM | Codice identificativo famiglia; SESSO | Sesso; CITTAD | Cittadinanza; STACIV | Stato civile; REL_PAR | Relazione di parentela; EST_NAS | Codice Istat Stato estero di nascita; ETA | Età puntuale; REDD_PENS_COMPL | Reddito da pensione ricalcolato; REDD_LAVDIP3 | Reddito da lavoro dipendente ricalcolato; REDD_CALCULO4 | Reddito complessivo ricalcolato; CL_REDD | Classificazione reddito; TIPO_SCUOLA | Tipo di scuola; TITSTU_RIC | Titolo di studio; FLAG_UNIVERSITARIO | Flag studente universitario; FLAG_PRECARIO | Flag precario nel mese di ottobre; FLAG_NOSOR_F | Flag che indica femmina in età attiva che non studia, non ha lavorato, senza reddito; FLAG_NOSOR_M | Flag che indica maschio in età

Figura 4.2 - Star schema semplificato del progetto SPoT-DeBEST



attiva che non studia, non ha lavorato, senza reddito; FLAG_OCC | Flag occupato nel mese di ottobre; LAV_ANNO_PREC | Flag che indica se ha lavorato almeno un giorno nell'anno; POS_ANNO | Intensità lavorativa annuale; OTT_ANNOCORR | Flag occupato nel mese di ottobre; CONDIZ_OTT_ANNOCORR | Condizione lavorativa principale ad ottobre; REDD_FAM | Reddito familiare; N_PERC_FAMIGLIA | Numero di percettori di reddito in famiglia; APPORTO_REDD_FAM | Percentuale di reddito individuale sul reddito complessivo familia-

re; **FAMIGLIE**: CODICE_FAM | Codice identificativo famiglia; COD_PROCOM | Codice Istat Provincia Comune di residenza; TIPFAM2 | Tipologia familiare; NCOMP | Numero componenti; SESSO_INT | Sesso dell'intestatario; NCOMP18_59 | Numero di componenti di età compresa tra i 18 e 59 anni; N_1839_NEE | Numero di componenti tra i 18 e i 39 anni che non hanno reddito né studiano; N_1839 | Numero di componenti tra 19 e 39 anni; N_18 | Numero di componenti under 18; N_CLETA_NEET | Numero di componenti tra i 15 e i 29 anni; N_ITA | Numero di componenti con cittadinanza italiana; N_STRA | Numero di componenti con cittadinanza straniera o apolide; CITTAD_FAM | Classificazione famiglia in base alla cittadinanza dei componenti; REDD_FAM | Somma del reddito dei componenti; PENS_FAM | Ammontare del reddito da pensione posseduto dalla famiglia; LAVDIP_FAM | Ammontare del reddito da lavoro dipendente posseduto dalla famiglia; N_CONTRIB_FAMIGLIA | Numero di contribuenti BDR; N_PERC_FAMIGLIA | Numero di percettori di reddito; N_NOSOR_F | Numero di persone in età attiva (15-64) che non studiano, non lavorano senza reddito di sesso femminile; N_NOSOR_M | Numero di persone in età attiva (15-64) che non studiano, non lavorano senza reddito di sesso maschile; OCCUPAZIONE_INT | Flag che indica se l'intestatario ha lavorato almeno un giorno nell'anno; N_STUD_PRIM | Numero di studenti della scuola primaria; N_STUD_SECI | Numero di studenti della scuola secondaria di I grado; N_STUD_SECH | Numero di studenti della scuola secondaria di II grado; N_UNII | Numero di studenti universitari di età inferiore ai 35 anni; N_UNI2 | Numero di studenti universitari di almeno 35 anni; NCOMPSTU18_24 | Numero di componenti di età compresa tra i 18 e 24 anni che studiano; DENOMINATORE | Numero di persone di età 18-59 (esclusi studenti 18-24); INT_LAV | Intensità lavorativa; N_STRANIERI | Numero di componenti con cittadinanza straniera o apolide; NCOMP_EQUIVALENTI | Denominatore usato per il calcolo del reddito equivalente; REDD_EQUIVALENTE | Reddito equivalente procapite familiare; TITSTU_CAPOFAMIGLIA | Titolo di studio del capofamiglia. Il capofamiglia è il percettore di reddito più alto all'interno della famiglia; nel caso in cui non ci siano percettori di reddito coincide con l'intestatario; APPORTO_REDD_INT | Apporto dell'intestatario al reddito familiare.

4.3.2 Il processo di Data warehousing end-to-end

Il processo implementato nell'ambito del progetto SPoT può essere schematizzato considerando il flusso informativo dei dati. La lettura dello schema, figura 4.3, è possibile sia secondo una prospettiva orizzontale, quindi tipicamente di fasi di processo, sia verticale in un'ottica di strumenti di esplorazione e analisi delle basi dati prodotte nelle differenti fasi di processo.

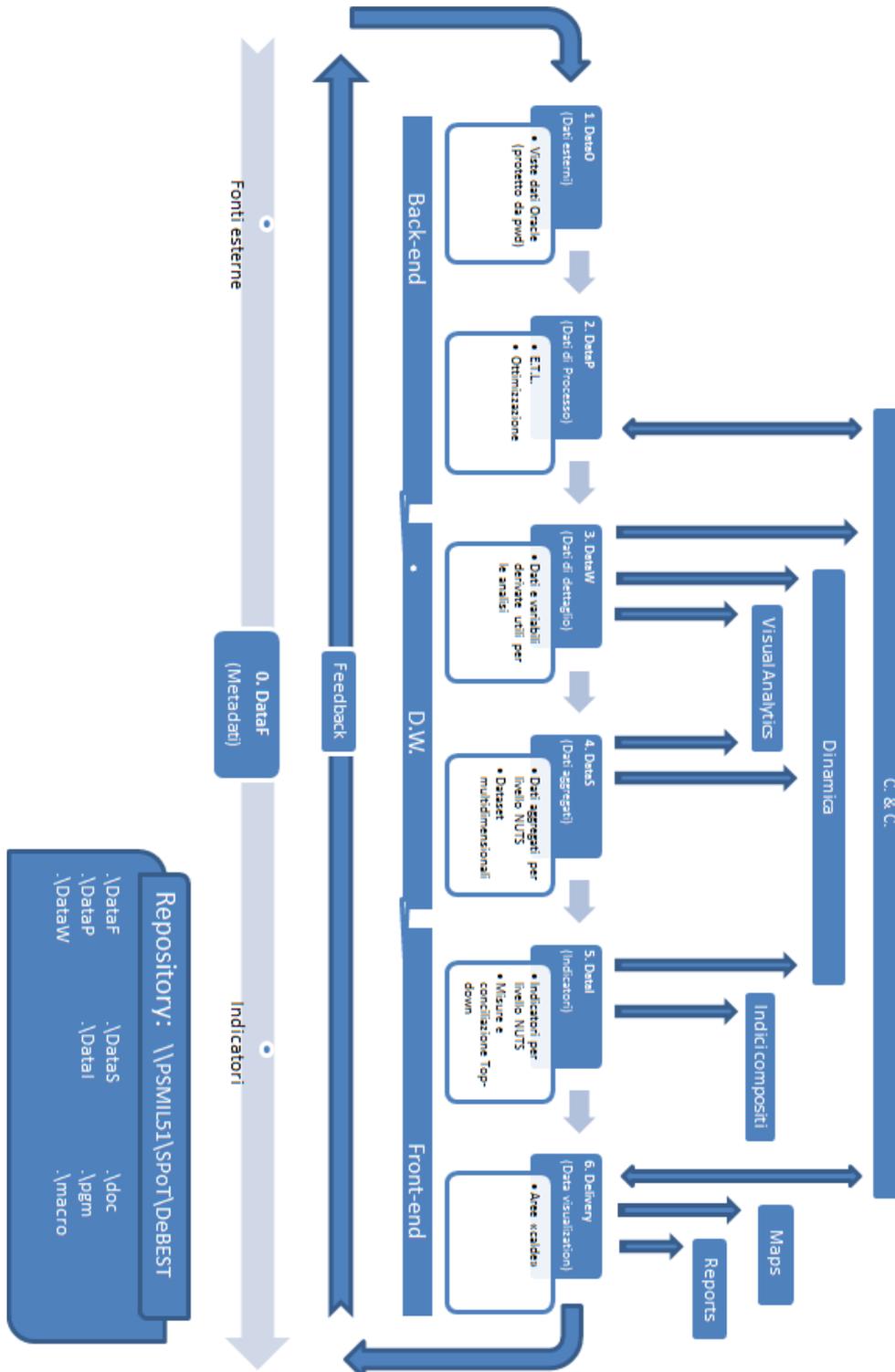
Rispetto ad uno sviluppo orizzontale, si parte dal *back-end* – basato sostanzialmente su “*viste*” Oracle (1. *DataO*) – e grazie a tecniche cosiddette di E.T.L. (*Extract, Transformation, Loading*) si accede alle basi dati di ArchIMEDe rese disponibili per la sperimentazione del progetto considerando, in questa prima fase progettuale, l'estrazione della totalità dei dati forniti. La trasformazione delle variabili si rende necessaria non solo per una omogeneità interna, ad esempio, delle nomenclature. Le procedure di trasformazioni sono d'obbligo nel passaggio da basi dati Oracle a quelle SAS in quanto – considerati come *DBMS* – differiscono per modalità d'archiviazione. In particolare, in SAS, il criterio della parsimonia viene adottato sia per le variabili alfanumeriche – considerando l'estensione massima delle modalità o la ricodifica – e, soprattutto, per quelle numeriche convertendo *flag* e classificazioni in variabili alfanumeriche.

Il risultato di questa attività di ottimizzazione confluisce in apposite tavole di lavoro intermedie (2. *DataP*). Successivamente viene predisposta l'area DW vera e propria (3. *DataW*) dove si trovano i “*fatti*” di interesse, in parte *tout court* riportati attraverso l'ETL delle variabili originariamente contenute nelle fonti accedute e in parte derivati grazie alla predisposizione di nuove variabili funzionali alle successive fasi di aggregazione e analisi dei dati.

Sono poi state predisposte, anche partendo da strategie *n-way*, basi dati aggregate (4. *DataS*) e di indicatori (5. *DataI*) rispetto alla gerarchia territoriale.

Quindi, nella fase successiva – *front-end* – vengono pubblicati gli indicatori attraverso strumenti di sintesi e di rappresentazione aggregata delle informazioni. Il *delivery* dei dati è volto a mettere in luce le informazioni rilevanti e le criticità su base territoriale a fini decisionali, e, in questo caso, a supporto delle politiche locali.

Figura 4.3 - Processo di *Data warehousing end-to-end* del progetto SPoT-DeBEST



Note: SPoT, attraverso il processo di *datawarehousing* qui sintetizzato, ha consentito l'avvio e la messa a disposizione dei dati in formato SAS anche per i progetti ISTAT "Misure di benessere e programmazione a livello comunale" e "Sistemi informativi sanitari regionali e indicatori socio-economici: sinergie dalla valorizzazione e integrazione di "nuove" fonti informative ufficiali".

Infine, il *feedback*, che può assumere diverse forme, è funzionale anche al miglioramento dei processi laddove, ad esempio, si osservi una carenza informativa o una problematica con caratteri di sistematicità.

Leggendo lo schema secondo una prospettiva verticale, in un'ottica di strumenti di esplorazione e analisi delle basi dati, gli obiettivi di studio cambiano a seconda delle differenti fasi di processo considerate. In particolare, gli *output* di tutte le fasi di processo possono essere di interesse per il controllo e la correzione (C&C) delle basi di dati e la validazione di variabili e indicatori. Più ristretti gli ambiti applicativi dei *visual analytics* che, pur consentendo lo sfruttamento di dati aggregati, nascono per l'analisi di specifici *data mart* o del DW intero in quanto permettono analisi esplorative cosiddette *in-memory*. Strutture multidimensionali (*MDDB*) garantiscono l'esplorazione di basi dati aggregate anche attraverso tradizionali strumenti di esplorazione interattiva di tipo *OLAP (On-Line Analytical Processing)*. Raccolte di indicatori derivati dai dati analitici possono essere utilizzati per la costruzione di indici sintetici (Cfr. Capitolo 5). Report e altri strumenti di sintesi e rappresentazione come le cartografie (p.es. *box plot map*, *LISA map*, *SEBS map* – Cfr. Capitolo 6) possono essere utilizzati in primis con finalità di supporto alle politiche locali.

4.3.3 Il repository

Per la condivisione degli aggregati prodotti nell'ambito del processo di *data warehousing* è necessario individuare un *repository*. Nel progetto Spot è stato scelto di realizzare un'area dati condivisa nella *intranet* d'Istituto. Infatti, nonostante il *repository* sia stato realizzato su un *file server* localizzato a Milano, nella sede territoriale della Lombardia, sono state poi predisposte le necessarie autorizzazioni – in sola lettura – al fine di consentire l'accesso da tutte le sedi dell'Istat (Figura 4.3).

4.4 Elementi di conclusione e prospettive future

L'utilizzo di metodi di gestione e organizzazione dei dati consoni a facilitare il processo decisionale basato sull'informazione rappresenta un orientamento sempre più d'interesse per i governi locali. In particolare, in questa sede, l'utilizzo del modello dati basato su fatti/dimensioni, tipico nell'ambiente di sviluppo di un DW, ha consentito più agevolmente la costruzione di indicatori e statistiche per l'individuazione di “*aree calde*” di ipotetico interesse per le politiche.

Di grande interesse per le prospettive di ricerca e studio dei fenomeni sono risultati gli applicativi di *visual analytics*³. Questi nuovi strumenti, in parte eredità di esperienze passate (p.es. *SAS Insight*), permettono infatti analisi esplorative *in-memory* per la rappresentazione di fenomeni analitici complessi, come ad esempio l'osservazione, attraverso l'uso di specifiche applicazioni, della dinamica scuola/lavoro/pensione. In questo ambito, prime interessanti sperimentazioni sono state condotte con strumenti di rappresentazione quali, ad esempio, il *Sankey diagram*, un particolare tipo di diagramma di flusso in cui l'ampiezza delle frecce è dimensionata in maniera proporzionale alla quantità di flusso.

³ Sperimentazione SPoT con SAS *visual analytics* – Sas Institute, Milano 21 febbraio 2017.

CAPITOLO V

COSTRUIRE GLI INDICI SINTETICI

di Matteo Mazziotta

5.1 Introduzione

In anni passati, molti fenomeni sociali ed economici, come lo sviluppo, la povertà, la qualità della vita, la disuguaglianza, l'infrastrutturazione ecc. erano misurati essenzialmente con un approccio unidimensionale, cioè ricorrendo a un unico indicatore. La recente letteratura tende a considerare questi fenomeni come complessi o multidimensionali, poiché sono definiti e rappresentati da un insieme di indicatori elementari. La misurazione di tali fenomeni complessi è un'operazione difficile e delicata, perché richiede semplificazioni che sono di per sé piuttosto soggettive, è sempre limitata da risorse e tempi ristretti e comporta, inevitabilmente, il rispetto di requisiti in competizione e conflitto tra loro, oltre a basarsi spesso su considerazioni pragmatiche. Tuttavia, è possibile associare, in modo coerente, la selezione degli indicatori elementari in grado di rappresentare il fenomeno considerato, alla scelta della "migliore" funzione di aggregazione, in modo da non perdere molta informazione statistica.

Lo scopo di questo capitolo è di presentare un indice composito generalizzato denominato MPI (Mazziotta-Pareto Index), adatto al caso in cui le componenti sono non sostituibili, ossia hanno tutte lo stesso peso (importanza) e non è ammessa una compensazione tra di esse (Munda e Nardo, 2005).

La letteratura scientifica riporta molti studi di importanti autori concernenti l'impiego di indici compositi per misurare fenomeni economici e sociali complessi, riguardanti aree geografiche. I problemi principali in questo approccio sono legati alla scelta del quadro teorico, alla disponibilità dei dati, alla selezione degli indicatori più rappresentativi e al loro trattamento, al fine di confrontarli e aggregarli.

Le fasi di costruzione di un indice composito possono essere riassunte come segue:

- *Definire il fenomeno da misurare.* La definizione del concetto deve trasmettere una chiara idea di ciò che si intende misurare con l'indice composito. In particolare, occorre fare riferimento a un quadro teorico che colleghi diversi sottogruppi e indicatori sottostanti.
- *Selezionare un insieme di indicatori elementari.* In teoria, gli indicatori devono essere scelti in base a rilevanza, accuratezza, tempestività, accessibilità, comparabilità e coerenza (Oecd, 2008). La fase di selezione è il risultato di un compromesso tra il rischio di possibili ridondanze dovute a sovrapposizioni di informazioni e il rischio di perdere informazioni.
- *Normalizzare gli indicatori elementari.* Questa fase mira a rendere confrontabili gli indicatori elementari. La normalizzazione deve essere effettuata prima di procedere all'aggregazione dei dati, dato che gli indicatori hanno spesso unità di misura diverse. Pertanto, occorre ricondurre gli indicatori a uno stesso standard, convertendoli in numeri puri, adimensionali. Un'altra motivazione a supporto della normalizzazione risiede nel fatto che alcuni indicatori possono essere correlati positivamente ("polarità" positi-

va) con il fenomeno da misurare, mentre altri possono essere correlati negativamente (“polarità” negativa). Gli indicatori devono essere normalizzati in modo tale che l’aumento di un indicatore normalizzato corrisponda a un aumento dell’indice composito. Vi sono diversi metodi di normalizzazione dei dati, come la trasformazione in ranghi, la trasformazione *Min-Max*, gli *z-scores* (o standardizzazione) e la trasformazione mediante una “distanza” da un riferimento. Per assegnare la stessa “importanza” a ciascun indicatore elementare è necessario applicare un criterio di trasformazione che renda gli indicatori indipendenti sia dall’unità di misura che dalla variabilità (p.es. la standardizzazione).

- *Aggregare gli indicatori normalizzati.* L’aggregazione è la combinazione di tutte le componenti al fine di ottenere uno o più indici compositi (funzioni matematiche). Esistono diversi metodi di aggregazione: il più utilizzato è quello additivo, che varia dalla somma dei ranghi di ciascun indicatore alla media ponderata di valori trasformati degli indicatori originari. Spesso, si ricorre anche a tecniche multivariate, come l’Analisi delle componenti principali (Dunteman, 1989) e il metodo tassonomico di Wroclaw (Harbison *et al.*, 1970), che si basa sulle “distanze”.

Questo approccio, ovviamente, non è scevro da diversi problemi, come il reperimento dei dati, la perdita di informazioni e l’arbitrarietà del ricercatore per quanto concerne: (i) la scelta degli indicatori, (ii) la normalizzazione, (iii) l’aggregazione e la ponderazione. I vantaggi, tuttavia, sono indiscutibili e possono essere riassunti come segue: (a) misurazione sintetica del fenomeno; (b) disponibilità immediata; (c) semplificazione dell’analisi dei dati geografici.

Molti lavori e analisi si sono imposti alla critica e la comunità scientifica è giunta alla conclusione che è impossibile ottenere un “metodo perfetto” che risulti universalmente efficiente. Al contrario, dati gli obiettivi specifici e i dati dei lavori, è necessario individuare gradualmente il “metodo migliore” in termini di solidità, affidabilità e coerenza.

I metodi additivi per la costruzione di indici compositi si basano su requisiti e proprietà che spesso sono non desiderabili o difficili da soddisfare. Per esempio, si assume una completa sostituibilità tra le componenti dell’indice: un deficit in una dimensione può essere compensato da un surplus in un’altra. Tuttavia, di norma non è accettabile una completa compensabilità tra i singoli indicatori ed è richiesta una distribuzione “equilibrata” dei valori.

5.2 Metodi

Il metodo proposto intende fornire una misura composita di un insieme di indicatori che siano considerati “non sostituibili”, ovvero tutte le dimensioni del fenomeno devono essere “bilanciate” (Mazziotta e Pareto, 2011). L’indice MPI è concepito per soddisfare le seguenti proprietà: (i) normalizzazione degli indicatori con un criterio specifico che depura gli indicatori sia dall’unità di misura che dalla loro variabilità (Delvecchio, 1995); (ii) sintesi indipendente da una “unità ideale”, dato che un insieme di “valori ottimali” è arbitrario, non univoco e può variare nel tempo (Aureli e Cutillo, 1996); (iii) semplicità di calcolo; (iv) facilità di interpretazione.

Queste proprietà possono essere soddisfatte con il seguente approccio. Com’è noto, distribuzioni di indicatori diversi, misurati in modo diverso, possono essere confrontate mediante la trasformazione in scarti standardizzati. Pertanto, si conver-

tono gli indicatori elementari in modo che oscillino tutti entro la medesima scala, con media 100 e scostamento quadratico medio pari a 10: i valori così ottenuti saranno compresi all'incirca nell'intervallo 70-130¹. In questo tipo di normalizzazione, il "vettore ideale" è l'insieme di valori medi ed è facile identificare sia le unità al di sopra della media (valori superiori a 100), sia le unità al di sotto della stessa (valori inferiori a 100). Inoltre, la normalizzazione con scarti standardizzati consente di liberare gli indicatori dalla loro variabilità e di assegnare loro lo stesso peso.

In tale contesto, si introduce un coefficiente di penalità che dipende, per ciascuna unità, dalla variabilità degli indicatori rispetto al valor medio ("variabilità orizzontale"): tale variabilità è misurata attraverso il coefficiente di variazione. L'approccio proposto penalizza il punteggio di ogni unità (la media aritmetica dei valori standardizzati) con una quantità che è direttamente proporzionale alla "variabilità orizzontale". Lo scopo è di favorire le unità che, a parità di media, presentano un maggiore equilibrio tra i valori degli indicatori (Palazzi, 2004).

Il metodo consente di costruire una misura "robusta" e meno "sensibile" all'inclusione o all'esclusione di un singolo indicatore (Mazziotta et al., 2010).

5.2.1 MPI

Il metodo per il calcolo dell'indice MPI prevede due fasi principali: normalizzazione e aggregazione.

1) Normalizzazione

Sia $\mathbf{X}=\{x_{ij}\}$ una matrice di n righe (unità statistiche) ed m colonne (indicatori elementari) e siano M_{x_j} ed S_{x_j} la media e lo scarto quadratico medio del j -esimo indicatore:

$$M_{x_j} = \frac{\sum_{i=1}^n x_{ij}}{n}; \quad S_{x_j} = \sqrt{\frac{\sum_{i=1}^n (x_{ij} - M_{x_j})^2}{n}}$$

Si costruisce la matrice standardizzata $\mathbf{Z}=\{z_{ij}\}$ in cui:

$$z_{ij} = 100 \pm \frac{(x_{ij} - M_{x_j})}{S_{x_j}} 10 \quad (1)$$

dove il segno \pm indica la "polarità" del j -esimo indicatore, ovvero il segno della relazione tra il j -esimo indicatore e il fenomeno da misurare ("+" se l'indicatore elementare rappresenta una dimensione considerata positiva e "-" se rappresenta una dimensione considerata negativa).

2) Aggregazione

Sia cv_i il coefficiente di variazione dell' i -esima unità:

¹ In base al teorema di Bienaymé-Cebycev, i valori della distribuzione compresi nell'intervallo 70-130 sono almeno l'89 per cento di tutti i valori.

$$cv_i = \frac{S_{z_i}}{M_{z_i}}$$

dove:

$$M_{z_i} = \frac{\sum_{j=1}^m z_{ij}}{m}; \quad S_{z_i} = \sqrt{\frac{\sum_{j=1}^m (z_{ij} - M_{z_i})^2}{m}}.$$

La forma generalizzata² dell'indice MPI può essere scritta nel seguente modo:

$$MPI_i^{+/-} = M_{z_i} \pm S_{z_i} cv_i \quad (2)$$

dove il prodotto $S_{z_i} cv_i$ è la "penalità" e il segno \pm dipende dal tipo di fenomeno da misurare.

Se l'indice composito è di tipo "crescente" o "positivo", ossia se a valori crescenti dell'indice corrispondono variazioni positive del fenomeno (p.es. lo sviluppo socio-economico), si utilizza la versione con penalità negativa MPI-. Viceversa, se l'indice è di tipo "decescente" o "negativo", ossia se a variazioni crescenti dell'indice corrispondono variazioni negative del fenomeno (p.es. povertà), si ricorre alla versione con penalità positiva MPI+.

5.2.2 AMPI

Nel caso si voglia adottare una metodologia di sintesi che tenga in considerazione i confronti nel tempo si può applicare una variante del Mazziotta-Pareto Index, denominata AMPI (Mazziotta e Pareto, 2016). Si tratta di una funzione per la sintesi di un insieme di indicatori elementari, nell'ipotesi che ciascuna componente non sia sostituibile con le altre (o lo sia solo in parte) e abbiano tutte la stessa importanza (De Muro et al., 2011). Tale approccio, detto anche non compensativo, richiede una distribuzione bilanciata di tutte le componenti elementari.

Il metodo di normalizzazione dei dati, al fine di consentire confronti assoluti nel tempo, si basa su un *re-scaling* degli indicatori elementari rispetto a due *goal-posts*, ovvero un minimo e un massimo che rappresentano il campo di variazione di ciascun indicatore per tutto il periodo considerato.

Si riportano di seguito i passi necessari per il calcolo dell'AMPI (Mazziotta e Pareto, 2016).

Data la matrice $\mathbf{X} = \{x_{ijt}\}$ con n righe (regioni), m colonne (indicatori) e p strati (anni), si calcola la matrice normalizzata $\mathbf{R} = \{r_{ijt}\}$:

$$r_{ijt} = \frac{(x_{ijt} - \text{Min}_{x_j})}{(\text{Max}_{x_j} - \text{Min}_{x_j})} 60 + 70 \quad (3)$$

² Si tratta di una forma generalizzata in quanto include "due indici in uno".

dove x_{ijt} è il valore dell'indicatore j nell'unità i per l'anno t e Min_{x_j} e Max_{x_j} sono i *goalposts* dell'indicatore j . I *goalposts* sono fissati in modo da porre uguale a 100 il totale Italia per l'anno base³, mediante le seguenti formule:

$$\text{Min}_{x_j} = \text{Rif}_{x_j} - \Delta_{x_j}$$

$$\text{Max}_{x_j} = \text{Rif}_{x_j} + \Delta_{x_j}$$

dove Rif_{x_j} è il valore di riferimento per l'indicatore j e $\Delta_{x_j} = (\text{Sup}_{x_j} - \text{Inf}_{x_j})/2$, indicando con $\text{Inf}_{x_j} = \min_{it}(x_{ijt})$ e $\text{Sup}_{x_j} = \max_{it}(x_{ijt})$.

Se l'indicatore j ha polarità negativa⁴, si calcola il complemento a 200 della (3)⁵.

Indicando con M_{r_i} e S_{r_i} , rispettivamente, la media e la deviazione standard dei valori normalizzati dell'unità i , l'indice composito è dato dalla formula:

$$\text{AMPI}_i^{+-} = M_{r_i} - S_{r_i} \text{cv}_{r_i} \quad (4)$$

dove $\text{cv}_{r_i} = S_{r_i}/M_{r_i}$ è il coefficiente di variazione dell'unità i .

Questo approccio è caratterizzato dall'uso di una funzione (il prodotto $S_{r_i} \text{cv}_{r_i}$) che consente di penalizzare le unità con valori sbilanciati degli indicatori normalizzati. La penalità si basa sul coefficiente di variazione ed è nulla se tutti i valori sono uguali. Lo scopo è di favorire le unità che, a parità di valor medio, hanno un maggior bilanciamento tra i vari indicatori.

Pertanto, l'AMPI è caratterizzato dalla combinazione di un "effetto medio" (M_{r_i}) e un "effetto penalità" ($S_{r_i} \text{cv}_{r_i}$) e indica come si colloca ciascuna unità rispetto ai *goalposts*.

³ In tal modo, le regioni con un livello complessivo degli indicatori superiore al valore Italia nell'anno base, avranno dei punteggi maggiori di 100; mentre quelle con un livello complessivo inferiore, avranno dei punteggi minori di 100.

⁴ La "polarità" di un indicatore elementare è il segno della relazione tra l'indicatore e il fenomeno da misurare ("+" se l'indicatore rappresenta una dimensione considerata positiva e "-" se rappresenta una dimensione considerata negativa).

⁵ I valori normalizzati saranno compresi, all'incirca, nell'intervallo (70; 130).

CAPITOLO VI

RAPPRESENTARE LE INFORMAZIONI TERRITORIALI

di GdL Istat SPoT¹

6.1 Introduzione

Il dato amministrativo non è immediatamente utilizzabile con le finalità della statistica ufficiale e anche considerandolo in un'ottica di statistiche sperimentali, ci sono limiti alla sua applicazione. Allo stesso tempo, però, rappresenta un patrimonio informativo inestimabile, soprattutto per la conoscenza dei fenomeni a livello microterritoriale. In particolare, le fonti amministrative contengono la quasi totalità dei dati funzionali a supportare le politiche locali e il loro monitoraggio (i.e. dati aventi le seguenti caratteristiche: periodicità almeno annuale, aree amministrative sub-regionali ed economicità).

La rappresentazione con indicatori di sintesi di fenomeni, siano essi sociali o economici, è sempre affetta da errori indipendentemente dalle modalità di raccolta dati. Nelle rilevazioni dirette esiste una vasta letteratura ed esistono approcci consolidati che permettono di stimare parte dell'errore (quello campionario) e di controllare la qualità del dato in termini di riduzione dell'errore ex-ante e di valutazione ex-post della sua entità con specifici indicatori. Nelle statistiche da fonte amministrativa, invece, le prassi sono ancora in fase sperimentale e/o in fase di definizione. Ciò induce a diffondere e utilizzare con cautela questo tipo di dato. Pertanto si può ragionare in termini di segnali confirmatori per l'individuazione di aree di interesse, previa un'accurata analisi critica della congruità delle definizioni operative, della correttezza degli algoritmi utilizzati nelle fonti amministrative e del rispetto dei requisiti di qualità dei dati ad uso statistico nei diversi domini territoriali, a garanzia dell'effettiva comparabilità delle variabili e degli indicatori da esse derivati. Più nello specifico, bisogna accertarsi che siano individuati e indicati dei chiari riferimenti temporali validi per tutte le realtà territoriali poste a confronto e che siano seguiti gli stessi criteri durante l'estrazione dei dati e il calcolo degli indicatori.

La cartografia è uno strumento adatto nella rappresentazione dei dati da fonti amministrative, purché le riflessioni critiche sulle fonti di errore non campionario, di cui sopra, non siano tali da portare a considerare le fonti amministrative in esame prive dei requisiti minimi per un loro utilizzo a fini statistici. Nel quadro più ampio di fornire un supporto statistico alle politiche locali si è cercato di individuare modalità di rappresentazione dei dati di fonte amministrativa che, pur mantenendo un elevato contenuto informativo, garantissero risultati coerenti con le statistiche ufficiali. Questo obiettivo si è tradotto nell'individuazione di strumenti di rappresentazione tali da non obbligare a rappresentare dati puntuali per le singole unità

¹ Gruppo di Lavoro Istat afferente alla linea di attività *SPoT - dati e metodologie per lo sviluppo delle Statistiche per le Politiche Territoriali* (Cfr. Istat, OdS DIRM 141/17): Ballabio Simona, Carbonara Monica, Cariello Salvatore, Carra Arianna, Carulli Assunta Lisa, Casacci Sara, Comune Maria Elena, Della Queva Stefania, Ferrazza Daniela, Giangregorio Monica, Mazziotto Matteo, Nicosia Manuela, Plescia Emanuela, Sicuro Lorella, Stoppio Sabrina, Succi Raffaella, Tucci Domenico, Vannoni Francesca, Verrecchia Flavio, Vitalini Alberto, Viviano Lorena Carmen Maria.

(p.es. valori assoluti o percentuali) e nemmeno statistiche sulla distribuzione del fenomeno (p.es. minimo, massimo, ecc.).

L'utilizzo delle cartografie tematiche per la distribuzione territoriale di un fenomeno in termini d'intensità prevede diversi ambiti di applicazione d'interesse:

- controllo e correzione a supporto dei ricercatori con competenze territoriali per l'identificazione di criticità nei dati dovute ad eventuali carenze dell'informazione secondaria (p.es. copertura);
- analisi e conoscenza del fenomeno per identificare aree in cui un fenomeno oggetto di studio è più "caldo" (i.e. critico).

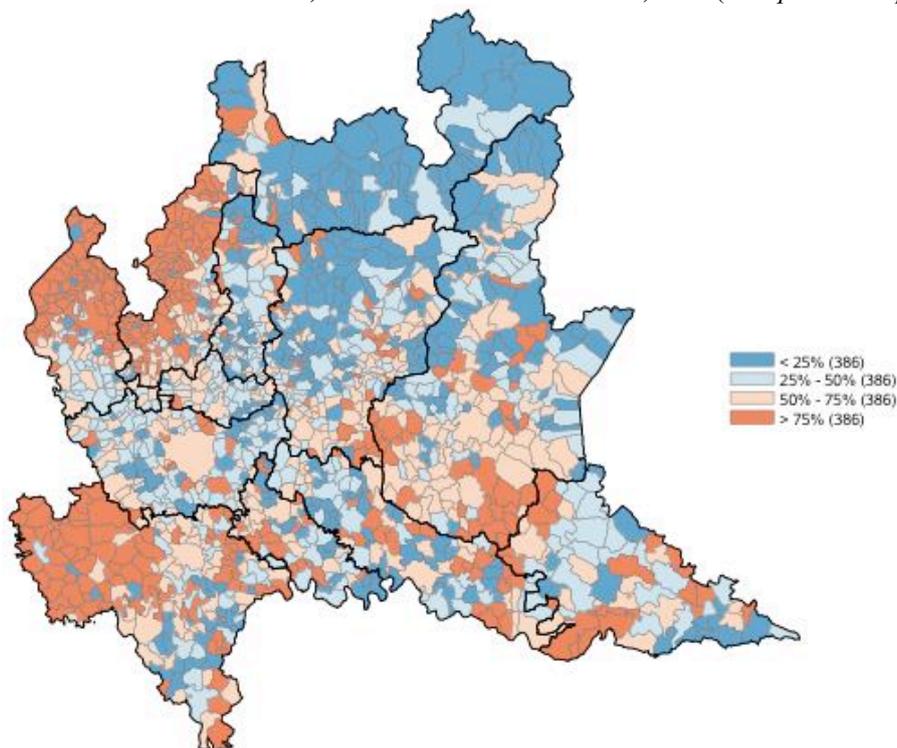
Per lo studio della distribuzione territoriale dell'incidenza di un fenomeno sono stati individuati quattro strumenti – che rispondono a differenti domande conoscitive specificate nelle relative schede – ritenuti particolarmente adatti per rappresentare l'intensità e la modalità di concentrazione o dispersione di un fenomeno a livello territoriale: *Quartile map*; *Box plot map*; *LISA Cluster map*; *SEBS box plot map*.

L'indicatore utilizzato nelle applicazioni è quello dei Neet – l'incidenza di giovani di età compresa tra i 15 e i 29 anni che non studia e non lavora (Cfr. Capitolo XIII).

6.2 Quartile map

L'utilizzo di cartografie tematiche con valori grezzi di un indicatore/indice raggruppati in quartili è la soluzione comunemente adottata per rappresentare la distribuzione territoriale di un fenomeno (Figura 6.1). Ad esempio, con la *temperature map*, ad una maggiore intensità dell'indicatore di vulnerabilità considerato corrisponderanno tonalità via via più "calde" dei territori rappresentati.

Figura 6.1 - Incidenza dei Neet, focus comunale. Lombardia, 2013 (raw quartile map)



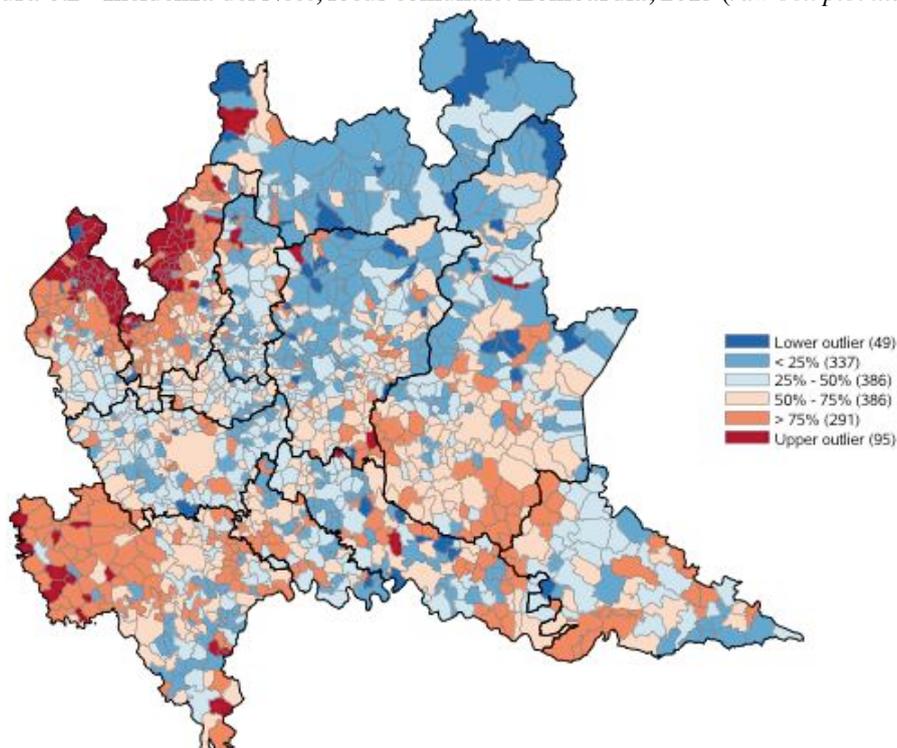
Scheda 6.1 - Quartile map

| | |
|----------------------------|--|
| Domanda conoscitiva | Come si distribuisce il fenomeno sul territorio? |
| Punti di forza | Consente di osservare in modo semplice e immediato, anche da non esperti, la distribuzione di un fenomeno. Essa, infatti, permette di esaminare la distribuzione territoriale di un indicatore e di individuare aree a maggiore o minore intensità del fenomeno studiato. |
| Punti di debolezza | <ul style="list-style-type: none"> ▪ Può fornire informazioni non consistenti su alcuni comuni, specialmente quelli con una limitata ampiezza demografica. ▪ Non permette confronti tra distribuzioni differenti riferite ad anni successivi e/o ad aree geografiche considerate separatamente (p.es. regioni). ▪ Non permette la corretta rappresentazione di indicatori con una distribuzione dei valori con valori uguali o molti vicini. ▪ Non evidenzia i valori estremi. ▪ Non è in grado di quantificare la presenza di autocorrelazione spaziale, cioè il grado di dipendenza di un valore assunto da una variabile in un'unità areale dal valore che la medesima assume in unità areali contigue. Molti indicatori sociali come il numero di residenti, il reddito pro-capite, i tassi di scolarizzazione non sono spazialmente indipendenti. Non si può peraltro escludere che la concentrazione di aree simili sia dovuta al caso o sia il frutto di un effetto grafico. |
| Soluzioni | <ul style="list-style-type: none"> ▪ È possibile ovviare al problema di fornire informazioni non consistenti su alcuni comuni, specialmente quelli con una limitata ampiezza demografica calcolando, se si hanno a disposizione dati relativi a più anni, il valore delle medie mobili. ▪ Per i confronti nel tempo e aree geografiche diverse si potrebbero definire soglie fisse basate su riflessioni ad hoc per ogni indicatore considerato (salvo conflitto con i vincoli prefissati). Si possono calcolare le variazioni sull'indicatore grezzo e mostrare i quartili delle variazioni concentrandosi sulle aree che diventano più "calde" (Cfr. scheda 6.5). ▪ Per tenere in considerazione la presenza di autocorrelazione spaziale è possibile usare le <i>LISA Cluster map</i> (Cfr. scheda 6.6). ▪ Per evidenziare i valori estremi, che spesso sono di stimolo alla riflessione, si può utilizzare la <i>Box plot map</i> (Cfr. scheda 6.2). |
| Usabilità | Nessun limite. |

6.3 Box plot map**6.3.1 Box plot raw map**

La *box plot map* consiste in una modifica di una cartografia in quartili, nella quale le aree, considerate *outlier*, del primo e quarto quartile sono separate. In particolare nella cartografia proposta (Figura 6.2) i colori arancione e rosso indicano zone "calde" (i.e. ad alta incidenza). Il rosso indica i comuni i cui valori sono maggiori del terzo quartile più 1,5 volte la distanza interquartile. L'arancione indica i restanti comuni che appartengono all'ultimo quartile della distribuzione dell'indice considerato. Il blu indica i comuni i cui valori sono minori del valore del primo quartile meno 1,5 volte la distanza interquartile. L'azzurro indica i restanti comuni che appartengono al primo quartile della distribuzione dell'indice considerato.

Figura 6.2 - Incidenza dei Neet, focus comunale. Lombardia, 2013 (raw box plot map)



Scheda 6.2 - Box plot map

| | |
|----------------------------|---|
| Domanda conoscitiva | Quali sono le aree di maggiore criticità? |
| Punti di forza | <ul style="list-style-type: none"> Consente di osservare in modo immediato la distribuzione di un fenomeno. Infatti, permette di esaminare la distribuzione territoriale di un indicatore e di individuare aree con differente intensità. Mette in evidenza le aree con valori estremi. |
| Punti di debolezza | <ul style="list-style-type: none"> Può fornire informazioni non consistenti su alcuni comuni, specialmente quelli con una limitata ampiezza demografica. Non permette confronti tra distribuzioni differenti riferite ad anni successivi e/o ad aree geografiche diverse (p.es. regioni). Non permette la corretta rappresentazione di indicatori con una distribuzione dei valori che presentino valori uguali o molti vicini. Può fornire informazioni non corrette sulla criticità di alcune aree nel caso di carenze informative "strutturali" nei dati (p.es. i frontalieri). |
| Soluzioni | <ul style="list-style-type: none"> Per la non robustezza dell'informazione comunale, si potrebbero oscurare (o evidenziare) i comuni per cui il dato non è ritenuto attendibile. In alternativa si potrebbe, ove possibile, aggregare comuni territorialmente contigui. Per le carenze informative nei dati è opportuno fornire una documentazione il più possibile esaustiva, evidenziando il campo di osservazioni ed eventuali problemi di copertura. Può ovviare in parte a questo problema l'omissione di specifiche aree (Cfr. scheda 6.3) se le finalità di analisi lo consentono. Uso congiunto con strumenti quali le <i>Growth maps</i> (Cfr. scheda 6.5). |
| Usabilità | Nessun limite. |

6.3.2 Box plot "blanked" map

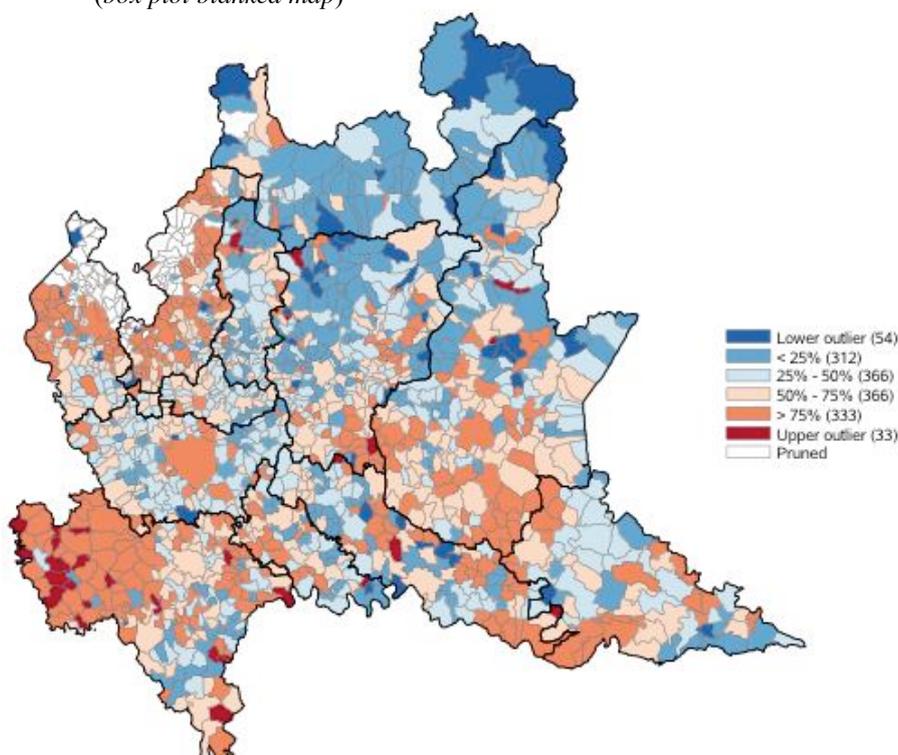
Aree «calde» (o «fredde») possono nascondere carenze informative dell'indicatore da fonte amministrativa. Ad esempio, nell'area insubrica parte della popolazione non studia o non lavora in Italia ma lo fa in Svizzera. Questa informazione è però assente nelle fonti considerate in ArchIMEDe. Pertanto, in figura 6.2,

queste aree si collocano in un'area *upper outlier*, che tuttavia potrebbe non corrispondere ad un'effettiva area di criticità. Nell'ipotesi di assenza di dati, a meno della conoscenza del fenomeno da parte del ricercatore, in figura 6.3 è stata esclusa (i.e. *pruned*) l'area «calda» delle provincie di Como, Varese e Sondrio.

Le aree per le quali la fonte è notoriamente non completa non sono considerate nelle analisi e pertanto sono rappresentate in bianco nella cartografica da cui il termine "*blanked*". Nel volume, si useranno anche altre strategie: omissione delle provincie con comuni frontaliere, evidenziazione dei comuni frontaliere così come previsto dalla legge (Cfr. Appendice 2).

Figura 6.3 - Incidenza dei Neet, focus comunale. Lombardia, 2013

(*box plot blanked map*)



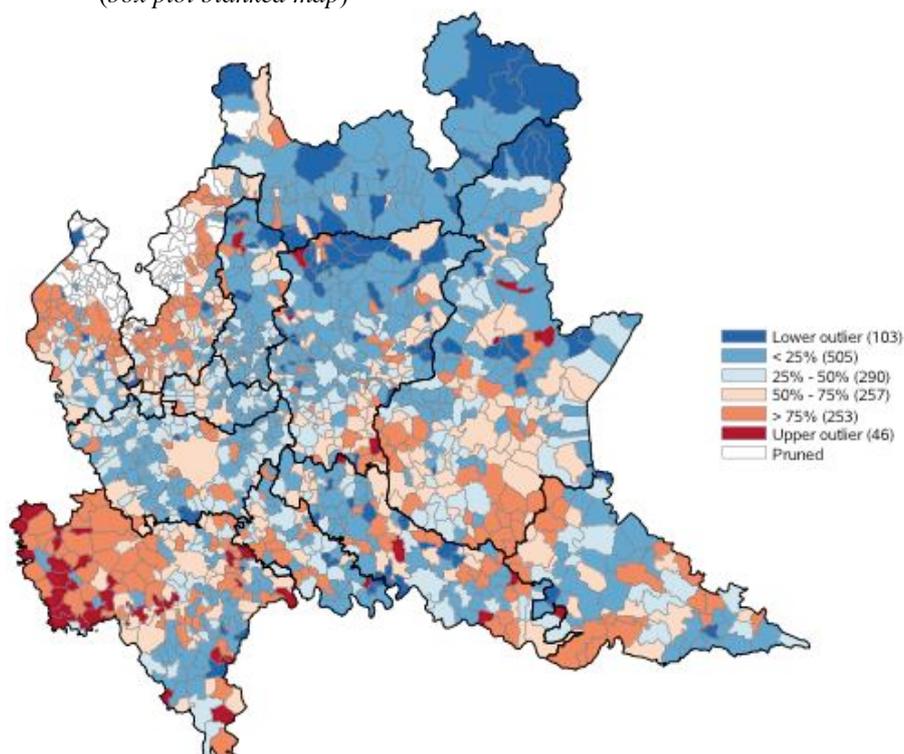
Scheda 6.3 - Box plot blanked map – focus comunale

| | |
|----------------------------|--|
| Domanda conoscitiva | Quali sono le aree di criticità effettiva? |
| Punti di forza | <ul style="list-style-type: none"> ▪ Si rimanda ai punti di forza elencati nella scheda 6.2; ▪ L'esclusione di alcune aree calde, sotto l'ipotesi di carenza informativa, evita di fornire informazioni errate su queste aree stesse. |
| Punti di debolezza | <ul style="list-style-type: none"> ▪ L'esclusione delle aree calde (nel caso presentato per Como, Varese e Sondrio) non esclude che il dato relativo ad aree (delle stesse provincie in primis, ma non solo) che non risultano calde, ma che ricadono nell'ultimo quartile, sia comunque inficiato dalle stesse carenze informative; ▪ L'assunzione dell'ipotesi di carenza informativa per specifiche aree. |
| Soluzioni | <ul style="list-style-type: none"> ▪ Integrare i dati mancanti con altre fonti; ▪ Stimare i dati mancanti; ▪ Valutare eventuali ulteriori aree di esclusione, facendo ricorso a conoscenze aggiuntive (p.es., per l'indicatore considerato, esistono norme, studi o ricerche sul fenomeno del frontalierato?). |
| Usabilità | Nessun limite. |

6.3.3 Box plot weighed map

Nelle precedenti rappresentazioni cartografiche, i comuni hanno lo stesso peso indipendentemente dall'ampiezza di specifiche popolazioni di riferimento. È però possibile applicare un peso, ad esempio proporzionale alla numerosità di una specifica popolazione. Nella cartografia successiva – figura 6.4 – i valori dei Neet sono stati pesati per la popolazione 15-29 anni.

Figura 6.4 - Incidenza dei Neet, focus popolazione 15-29 anni. Lombardia, 2013
(box plot blanked map)



Scheda 6.4 - Box plot blanked map – focus popolazione 15-29 anni

| | |
|----------------------------|---|
| Domanda conoscitiva | Quali popolazioni insistenti in aree di criticità effettiva? |
| Punti di forza | <ul style="list-style-type: none"> Si rimanda ai punti di forza elencati nella scheda 6.2. Consente di pesare i comuni rispetto alla popolazione target. Nell'esempio dei Neet si tiene conto della diversa distribuzione della popolazione 15-29 anni a livello comunale, compensando la numerosità dei comuni con più abitanti e quelli con meno abitanti e consentendo dei confronti del fenomeno più appropriati. |
| Punti di debolezza | <ul style="list-style-type: none"> Si rimanda ai punti di debolezza elencati nella scheda 6.3. |
| Soluzioni | <ul style="list-style-type: none"> Si rimanda alle soluzioni elencate nella scheda 6.3. Calcolare un modello di regressione tra una variabile indipendente costruita come incrocio fra l'ampiezza della popolazione dei singoli comuni (ricodificata p.es. in classi) e le singole province e il valore dell'indicatore come variabile dipendente. A questo punto, si possono utilizzare i residui standardizzati della regressione per le analisi (Becker, R.A., Denby, L., McGill, R., and Wilks, A.R.; 1987), "Analysis of data from the Places Rated Almanac," <i>The American Statistician</i>, 41(3), 169-186). Questo metodo essenzialmente usa la differenza fra valori grezzi e valori previsti dal modello per valutare quanto ogni comune si discosta dal valore previsto a partire da un determinato livello di popolazione in una provincia. |
| Usabilità | Utilizzo guidato. |

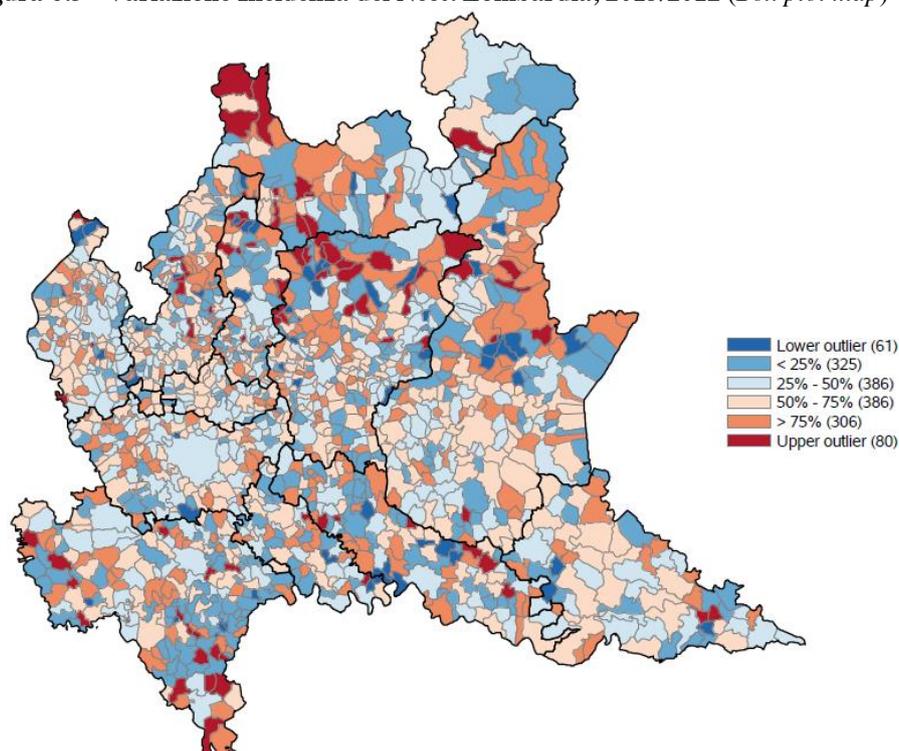
6.3.4 Box plot map: variazioni

Per valutare il cambiamento nel tempo si possono calcolare le variazioni sull'indicatore grezzo e mostrare una *Box plot map* delle variazioni concentrandosi sulle aree che diventano più calde (Figura 6.5).

Scheda 6.5 - Box plot map – focus temporale

| | |
|----------------------------|---|
| Domanda conoscitiva | Dove cresce la temperatura? |
| Punti di forza | <ul style="list-style-type: none"> ▪ Si rimanda ai punti di forza elencati nella scheda 6.2. ▪ Il calcolo dell'indicatore di variabilità non è influenzato da scelte soggettive del ricercatore. |
| Punti di debolezza | <ul style="list-style-type: none"> ▪ Si rimanda ai punti di debolezza elencati nella scheda 6.2. ▪ Può amplificare i problemi legati alla qualità legati agli archivi amministrativi (p.es. cambiamento del processo di raccolta del dato). |
| Soluzioni | <ul style="list-style-type: none"> ▪ Utilizzo congiunto con altri strumenti. |
| Usabilità | Nessun limite. |

Figura 6.5 - Variazione Incidenza dei Neet. Lombardia, 2013/2012 (*Box plot map*)



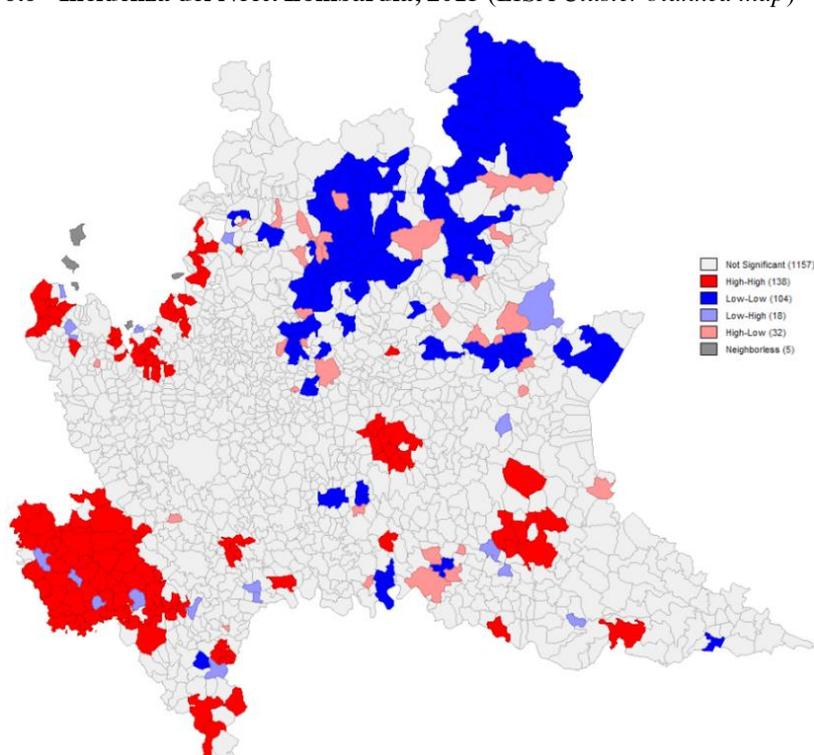
6.4 LISA cluster map

Sono state introdotte in letteratura delle statistiche che tengono in considerazione l'autocorrelazione spaziale. Queste statistiche consentono di: quantificare la correlazione, individuare *cluster* di comuni e verificare che la correlazione non sia dovuta al caso. In questa sede si propone l'utilizzo dell'indice LISA – *local indicator spatial association* – proposto da Luc Anselin (1994). La cartografia tematica “LISA Cluster map” è il modo più intuitivo per rappresentare graficamente e contemporaneamente i valori Lisa statisticamente significativi e l'informazione circa l'ordine di grandezza del fenomeno del comune e dei suoi vicini. La *Cluster Map* è

una cartografia tematica che mostra solo i comuni con valori statisticamente significativi di LISA, classificati secondo cinque categorie (Figura 6.6):

- Non significativo (Aree che non sono significative al livello 0,05);
- *High-High* (Alto valore dell'indicatore e comuni vicini con alti valori dell'indicatore);
- *Low-Low* (Basso valore dell'indicatore e comuni vicini con bassi valori dell'indicatore);
- *Low-High* (Basso valore dell'indicatore e comuni vicini con alti valori dell'indicatore);
- *High-Low* (Alto valore dell'indicatore e comuni vicini con bassi valori dell'indicatore).

Figura 6.6 - Incidenza dei Neet. Lombardia, 2013 (LISA Cluster blanked map)



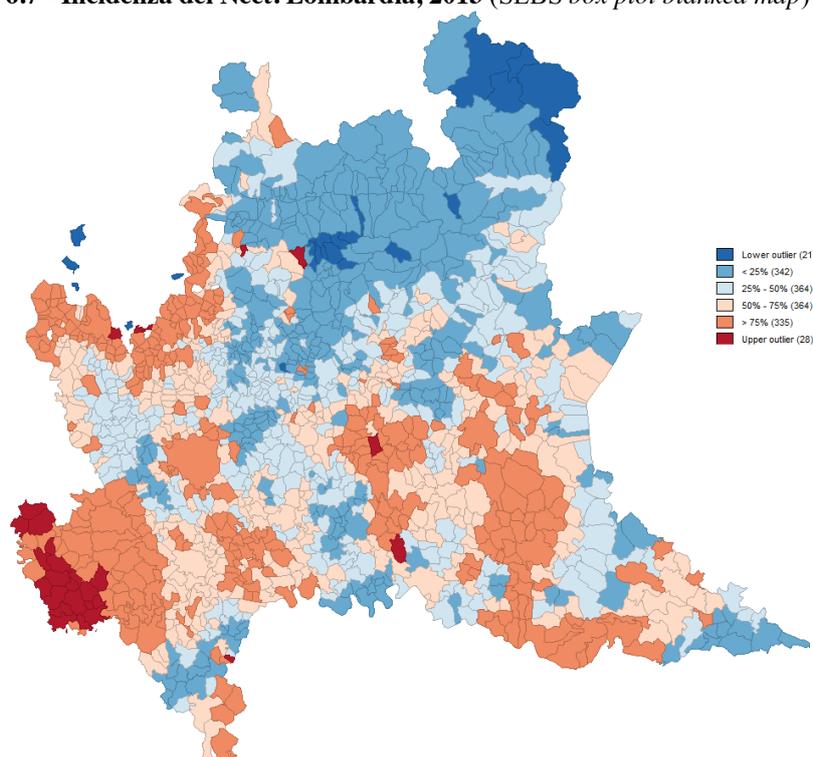
Scheda 6.6 - LISA Cluster blanked map

| | |
|----------------------------|---|
| Domanda conoscitiva | Quali cluster geografici segnalano indizi di criticità? |
| Punti di forza | <ul style="list-style-type: none"> ▪ Consente di visualizzare in maniera immediata l'associazione spaziale. ▪ Consente di valutare eventuali relazioni tra aree territoriali vicine e verificare che la correlazione del fenomeno in aree limitrofe non sia dovuta al caso. |
| Punti di debolezza | <ul style="list-style-type: none"> ▪ Se utilizzato da solo, questo strumento potrebbe essere di difficile comprensione per un pubblico non esperto o non specializzato; ▪ Il software utilizzato in questa applicazione non consente di fissare il "seme" iniziale. Questo fatto determina la produzione di cartografie differenti a partire dai medesimi dati. |
| Soluzioni | <ul style="list-style-type: none"> ▪ Fissare il <i>seed</i> iniziale utilizzando altri software; ▪ Accompagnare con altri strumenti. |
| Usabilità | Utilizzo prevalentemente interno o con documentazione a supporto. |

6.5 SEBS box plot map

I valori *spatial empirical bayesian smoothed rate (SEBS)*, calcolati per ogni area, sono rappresentati attraverso la cartografia tematica *box plot*. I SEBS sono un miglioramento della tecnica degli *empirical bayesian smoothed rate (EBS)* (Devine e Louis, 1994; Anselin, 2004), pensata per tenere sotto controllo la non omogeneità spaziale delle varianze. Consiste nel calcolare, per ogni area (p.es. comunale), un valore dell'incidenza dei Neet dato dalla media ponderata tra i valori dell'incidenza del comune e la media regionale, con pesi proporzionali alla popolazione a rischio. In altre parole nei comuni con una limitata ampiezza demografica il valore comunale grezzo dell'incidenza dei Neet sarà modificato in modo marcato e "trascinato" (*shrinkage*) verso il valore della media regionale, mentre negli altri comuni il valore grezzo dell'incidenza dei Neet sarà modificato poco o per niente. Nel calcolo dei SEBS, al posto della media regionale, si utilizza come fattore di *shrinkage* la media dei valori dell'incidenza dei Neet rilevata nei comuni vicini, assumendo che la dipendenza di un valore assunto da una variabile in un'unità areale, rispetto al valore che la medesima assume in unità areali contigue, sia più adatta a rilevare l'autocorrelazione spaziale che caratterizza il fenomeno e a mostrare i pattern territoriali (Figura 6.7). Il concetto di vicinanza è operazionalizzato da una matrice di contiguità.

Figura 6.7 - Incidenza dei Neet. Lombardia, 2013 (SEBS box plot blanked map)



Scheda 6.7 - SEBS box plot blanked map

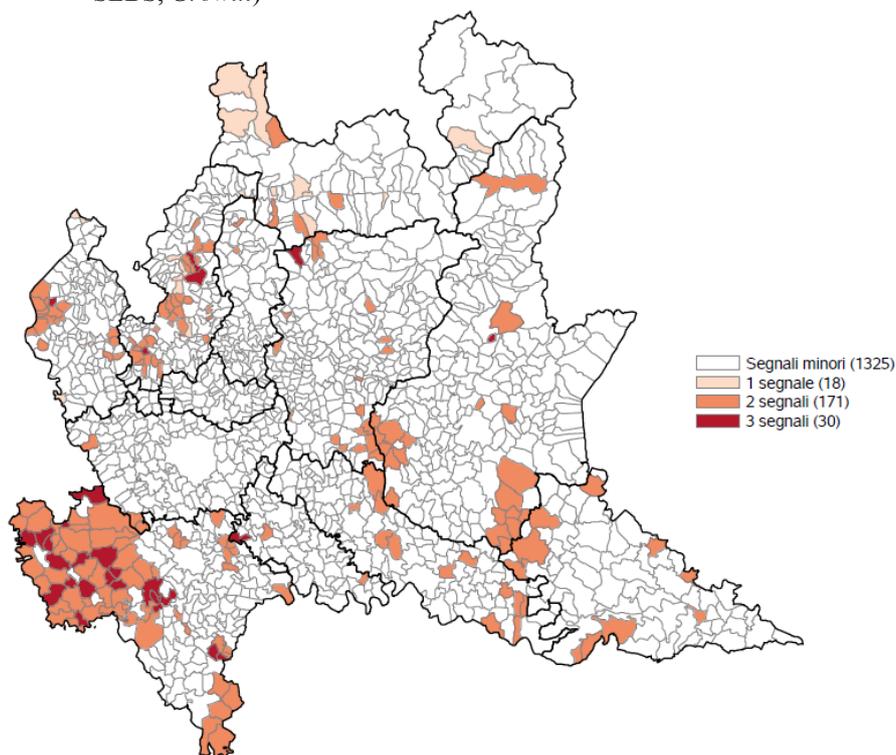
| Domanda conoscitiva | Quali aree contigue segnalano indizi di criticità? |
|---------------------------|--|
| Punti di forza | <ul style="list-style-type: none"> ▪ Tiene sotto controllo i valori dell'indicatore non consistenti, ovvia al problema di fornire dati non affidabili sui comuni di limitata ampiezza demografica. |
| Punti di debolezza | <ul style="list-style-type: none"> ▪ Potrebbe essere di difficile comprensione per utenti non specializzati; ▪ Non permette di studiare l'andamento del fenomeno nei singoli comuni; ▪ Assunzione dell'ipotesi di autocorrelazione spaziale del fenomeno. |
| Soluzioni | <ul style="list-style-type: none"> ▪ Togliere confini territoriali; ▪ Utilizzo congiunto con altri strumenti. |
| Usabilità | Utilizzo guidato. |

6.6 Cartografie miste

Si possono anche individuare aree critiche attraverso la sintesi di informazioni provenienti da più strumenti (Figura 6.8). L'idea alla base è quella di rafforzare gli indizi di criticità attraverso un approccio confermativo. L'ipotesi di lavoro qui proposta dipende anche dagli strumenti e dalle applicazioni considerate. In particolare:

- Un segnale solo se le variazioni sono *outlier* e se il comune appartiene alle provincie di Varese, Como e Sondrio (i.e. aree considerate con carezza informativa);
- Due segnali di tre tra: LISA *High-High* o *High-Low*; SEBS: quarto quartile; variazioni: quarto quartile;
- Tre segnali di tre: LISA *High-High* o *High-Low*; SEBS: quarto quartile; variazioni: quarto quartile.

Figura 6.8 - Aree di interesse per politiche locali sui Neet. Lombardia (Segnali LISA, SEBS, *Growth*)



Scheda 6.8 - *Mixed map*

| | |
|----------------------------|--|
| Domanda conoscitiva | In quali aree i diversi strumenti considerati segnalano contemporaneamente indizi di criticità? |
| Punti di forza | <ul style="list-style-type: none"> ▪ Fornisce un'informazione molto sintetica. ▪ Tiene conto di aspetti rilevanti in ottica sia sincronica sia diacronica. |
| Punti di debolezza | Perdita di informazione. |
| Soluzioni | Questo approccio potrebbe essere usato per dare una prima indicazione della situazione e per evidenziare aree di interesse per le politiche locali. |
| Usabilità | Utilizzo guidato con documentazione a supporto. |

INTRODUZIONE ALLA PARTE TERZA

APPLICAZIONI SU DATI AMMINISTRATIVI NEGLI ENTI LOCALI

di Antonio Lentini

La parte terza del volume ci porta ad esaminare alcune applicazioni concrete nell'utilizzo di ArchIMEDe da parte di enti appartenenti alla famiglia del Sistan¹. L'interesse di questo approdo è dato dal fatto che si esce, in questo caso, dai confini dello spazio della statistica inteso in senso stretto (che abbiamo visto nella prima sezione, parlando di fonti, nella seconda, parlando di metodi e che, infine, vedremo nella quarta, dove si trovano applicazioni nell'ambito dell'Istat stessa), per allargare lo sguardo alla rete di soggetti pubblici e privati che fornisce al paese e agli organismi internazionali l'informazione statistica ufficiale. È nell'ambito del Sistan che si muovono le applicazioni che vengono proposte in questa sezione: il loro interesse è dato dal fatto che nascono in risposta a questioni e problematiche che sorgono dal territorio sul quale incide anche la funzione statistica, incardinata in un caso in un comune e nell'altro in una regione.

È nel *policy cycle* che si collocano le analisi statistiche e l'integrazione tra basi dati garantita dalla disponibilità di ArchIMEDe che permette di provare a dare risposta ad alcune domande legate strettamente ai bisogni dei cittadini. Domande semplici ma che, come spesso accade, hanno componenti di multidimensionalità difficili da gestire e ancor prima, da comprendere, se non si hanno a disposizione strumenti di lettura e analisi che approssimino tale complessità. Questa particolarità si traduce poi anche in una sollecitazione ulteriore del sistema statistico che qui è chiamato a dare risposte almeno in tre momenti del processo decisionale pubblico: in fase di individuazione delle azioni da implementare (con la produzione di scenari alternativi, *what if*, proiezioni, ecc.) in fase di attuazione delle politiche (con il monitoraggio e, soprattutto, l'analisi del tasso di copertura) e infine in fase conclusiva (con gli studi sulla efficacia, intesa in senso ampio, delle *policy* messe in campo).

C'è però anche un altro aspetto che vale la pena di evidenziare: gli attori pubblici che fanno parte del Sistan sono *soggetti di governo del territorio* e sono essi stessi produttori di una quota di informazione che, seppur di natura amministrativa (quindi priva dei requisiti per poter essere annoverata tra le fonti statistiche) costituisce una ricchezza determinante per la conoscenza dei fenomeni². Si genera quindi in questi ambiti un punto di contatto vivo tra il dato statistico e l'informazione amministrativa. Nel seguito si darà conto di come l'incontro tra questi ambiti di conoscenza possa dare dei risultati interessanti e promettenti in termini di governo dei processi locali. Più in generale, va riconosciuto che anche nella pubblica amministrazione si sta affermando l'idea che non si possano fare *politiche* senza la disponibilità di dati statistici e analisi adeguate. Non è cosa banale in una amministrazione abituata ad una certa autoreferenzialità e dove il *numero* viene usato spesso per contare la quantità di atti amministrativi prodotti piuttosto che enumerare i soggetti coinvolti o l'impatto delle proprie politiche. E che spesso continua a confondere la cultura del dato come *misurazione della performance* con quella della valutazione come giudizio e, quindi, possibile punizione. Un ultimo aspetto che conviene citare è il fatto che ArchIMEDe, progetto lungimirante, è riuscito a portare a sistema anche dati che prima erano in ambiti diversi, superando la duplice barriera di tipo informatico, tra basi dati, e gestionale, dovuta all'appartenenza delle informazioni ad amministrazioni diverse.

Nei due capitoli che seguono questa breve introduzione, si vedrà l'uso di dati amministrativi delle regioni e comuni, in particolare per analizzare le politiche formative e sul mercato del

¹ "Che cosa è il Sistan" <http://www.sistan.it/index.php?id=422>.

² Sfida interessante rispetto alla trattabilità e assimilabilità con le fonti statistiche intese in senso stretto.

lavoro a livello regionale con la disponibilità dei dati degli universi di riferimento, così come per analizzare gli interventi sociali dei comuni. Ma sullo sfondo vi è anche l'ambizioso obiettivo di esplorare l'ambito sanitario partendo dalle condizioni socio-economiche delle popolazioni e sfruttando gli enormi archivi amministrativi e epidemiologici delle regioni. Più in particolare relativamente ai due capitoli inseriti in questa sezione, se ne può dare qui una breve descrizione per inquadrarli.

Il primo contributo, che riporta l'esperienza di Brescia nell'utilizzo di ArchIMEDe a supporto delle politiche locali, affronta temi eterogenei che tuttavia rappresentano esemplificazioni delle potenzialità dell'utilizzo di un sistema di microdati integrati di natura amministrativa nella pianificazione, programmazione e gestione dell'ente. In particolare:

- Il primo tema riguarda il livello e le caratteristiche differenziali della fecondità delle bresciane, e ha l'obiettivo di fornire al decisore un quadro di medio periodo utile per stimare l'andamento della fecondità, e a seguire della natalità, e quindi per pianificare i servizi che sono legati a quest'ultima.
- Il secondo tema – che si collega ad esigenze di programmazione dei servizi – riguarda le caratteristiche differenziali degli invalidi civili e soprattutto delle caratteristiche delle famiglie in cui gli stessi vivono. L'invalidità civile rappresenta un segmento di domanda contenuto nel numero, ma rilevante per le risorse che assorbe, e la disponibilità di dati sulle caratteristiche socio-economiche delle famiglie in cui vivono può aiutare a valutare le strategie di intervento pubblico, anche relativamente alla copertura dei costi dei servizi.
- Il terzo tema – che rimanda alla gestione – riguarda l'analisi degli aspetti distributivi delle politiche tariffarie anche in questo caso avendo attenzione alle famiglie. Le agevolazioni tariffarie comunali per i servizi a domanda individuale sono legate alla capacità contributiva delle famiglie stesse, costruita sull'indicatore Isee. L'indicatore tiene conto delle caratteristiche familiari solo in termini di numero e capacità di generare reddito, ma non di tipologia, e quindi di relazioni tra i componenti della famiglia.

Il secondo contributo, che riporta l'esperienza di Polis Lombardia, l'Istituto regionale per il supporto alle politiche della Lombardia (già Eupolis Lombardia), nasce dall'esigenza concreta di accompagnare la decisione annuale relativa alla programmazione dell'offerta formativa sul territorio regionale (da parte delle amministrazioni provinciali, che ne hanno delega) con una analisi che, senza certamente imporre obblighi, orienti le decisioni locali sulla base di informazioni provenienti dagli archivi amministrativi locali e dalle basi dati di ArchIMEDe. Questa esigenza, legata alla *governance* dei processi di formazione che la Regione esercita, è stata affrontata utilizzando anche dati amministrativi per costruire un vero e proprio modello (presentato nei suoi aspetti essenziali nel capitolo VIII) che sfrutta in particolare le comunicazioni obbligatorie per provare a stimare la domanda per figura professionale processando la serie degli avviamenti nei periodi precedenti. Unendo alle fonti regionali di carattere amministrativo quelle provenienti da fonti amministrative nazionali (residenza, istruzione, lavoro, reddito, via Miur e Inps per esempio), nell'ambito della sperimentazione del progetto ArchIMEDe in particolar modo rispetto a precarietà lavorativa e popolazione insistente, è stato possibile effettuare una seconda analisi che si è potuta spingere fino ad un livello territoriale molto fine, ovvero quello comunale: obiettivo, in questo caso, analizzare la sottopopolazione dei giovani senza segnali amministrativi di istruzione, formazione e lavoro evidenziandone le principali caratteristiche socio-demografiche e la durata della permanenza nella condizione di Neet.

I due esercizi sono solo degli esempi delle potenzialità che l'integrazione delle basi dati amministrative può offrire ed altri studi si possono già adesso disegnare in altri campi, come quello socio-sanitario, lavorando sia sul livello individuale, sia su quello familiare, per fare emergere situazioni di difficoltà là dove si vengono a sommare, nell'ambito di un nucleo.

CAPITOLO VII
L'ESPERIENZA DEL COMUNE DI BRESCIA
PER LE POLITICHE LOCALI¹
di Marco Trentini

7.1 Introduzione

L'art. 14 del Testo Unico Enti Locali (d.lgs. 267/2000) annovera la statistica tra le funzioni di governo delegate al Sindaco. Si tratta di una visione della statistica finalizzata alla tutela degli interessi informativi generali, che non evidenzia adeguatamente il fatto che la statistica comunale dovrebbe essere innanzitutto una funzione del governo locale. In quanto tale la statistica pubblica locale deve rispondere a due esigenze fondamentali: il supporto conoscitivo ai processi decisionali pubblici e la partecipazione dei cittadini alla crescita della comunità (attraverso la diffusione di informazioni sugli aspetti demografici, sociali, economici delle comunità stesse).

In questa capitolo trattiamo della funzione statistica a supporto dei processi decisionali pubblici a livello locale, lasciando ad altre occasioni il secondo e non meno rilevante tema della statistica come strumento di democrazia.

La funzione della statistica quale supporto a processi decisionali che hanno natura politica, per essere efficace (vale a dire offrire un prodotto che sia pertinente rispetto alle esigenze²) deve necessariamente tenere conto delle modalità in cui è strutturata l'attività delle istituzioni.

È ben noto che la vita politico-amministrativa locale si articola almeno su tre orizzonti temporali distinti: il primo è legato al ciclo elettorale del Sindaco e del Consiglio comunale e quindi alla pianificazione di medio termine dell'attività del comune; il secondo ai cicli triennali della programmazione economica e finanziaria; il terzo al ciclo annuale della programmazione e del controllo della gestione. Pianificazione, programmazione e gestione sono i tre livelli in cui si articolano i processi decisionali a livello locale, e di essi la statistica comunale pubblica deve tenere conto offrendo informazioni quantitative, ma non solo, pertinenti, tempestive, obiettive, corrette e via dicendo.

La funzione statistica non si riduce, in questa visione, ad effettuare elaborazioni di dati, ma ha un ruolo più ampio e significativo che comprende l'individuazione e selezione di modelli interpretativi della realtà (presi dalle discipline scientifiche come demografia, economia, sociologia, ecc.), la selezione delle fonti dei dati (esistenti o da costruire) e la scelta dei metodi di trattamento delle informazioni coerenti con i modelli e le fonti dei dati.

Uno dei problemi di maggiore rilievo che gli uffici di statistica devono affrontare per svolgere la propria funzione istituzionale è la disponibilità di dati a livello

¹ Il presente contributo riprende – sistematizzandola – una presentazione tenuta in occasione della Giornata della statistica del 22 ottobre 2015 a Cremona nel convegno “*Better data, better lives*, Strumenti di diffusione dell'informazione statistica nati dalla collaborazione fra enti” dal titolo *L'esperienza del Comune di Brescia come sperimentatore del progetto ArchIMEDe*. Grazie al prof. Luigi Riva per le critiche, i consigli e i suggerimenti che hanno notevolmente migliorato il testo.

² La pertinenza è uno di principi fondamentali a cui deve adeguarsi la statistica pubblica, tra cui imparzialità, obiettività, inattaccabilità metodologica, replicabilità, ... così come stabiliti a livello nazionale e internazionale.

di dettaglio subcomunale, di qualità e riferiti ai soggetti a cui è indirizzata prevalentemente l'attività comunale, quindi individui e famiglie.

Il sistema ArchIMEDe, ancorché sperimentale, ha le caratteristiche elencate in precedenza, e si presta quindi a diventare una fonte di dati utile a livello locale per la elaborazione di statistiche, studi e ricerche funzionali a supportare le amministrazioni nelle loro scelte.

L'ufficio di statistica del Comune di Brescia ha colto immediatamente la possibilità consentita da Istat di utilizzare i dati di ArchIMEDe, in specifico il rilascio sulla precarietà familiare, per sperimentare potenzialità e limiti di questa fonte per elaborazioni su temi di interesse locale. In questa sede presentiamo tre lavori commissionati da differenti assessorati del Comune, in cui sono stati utilizzati dati di ArchIMEDe – *Precarietà familiare*. I temi affrontati sono eterogenei e vengono presentati, in questa sede, come esempi delle potenzialità d'uso di un sistema di microdati come ArchIMEDe nella pianificazione, programmazione e gestione dell'Ente.

7.2. Il quadro concettuale

Senza entrare in un esame di dettaglio di un tema già presentato in un precedente lavoro (Riva e Trentini, 1998), crediamo sia utile riprendere sinteticamente alcuni aspetti del quadro concettuale di riferimento allo scopo di fornire il quadro concettuale delle presente nota.

Per offrire un supporto adeguato ai processi decisionali pubblici la statistica ufficiale deve essere in grado di attenersi alle modalità con cui la politica organizza le proprie attività. Il fattore critico è dato dai cicli politico-normativi che si articolano sui tre diversi orizzonti temporali già citati nell'introduzione.

La statistica opera fornendo il proprio supporto prevalentemente quantitativo per ognuno dei livelli vale a dire, per sintetizzare:

- Pianificazione intesa come supporto informativo al fine di ridurre i rischi in un contesto di ambiente mutevole, attraverso l'anticipazione di nuovi comportamenti, che potrebbero impattare sui servizi, e dei modelli gestionali dei servizi;
- Programmazione quale supporto informativo per la stima della domanda potenziale dei servizi, avendo come finalità la loro riprogettazione lungo le direttrici definite nella pianificazione;
- Gestione: la statistica offre un supporto informativo anche per fornire un criterio di definizione delle priorità di intervento soprattutto nell'ambito sociale, date le politiche di welfare.

Senza entrare in un esame di dettaglio dei vari livelli presentiamo di seguito alcuni schemi che descrivono in modo più dettagliato i tre livelli mettendo in evidenza da un lato il quadro di riferimento, dato l'orizzonte temporale, e dall'altro gli strumenti informativi funzionali a rispondere a quelle esigenze e le fonti dei dati.

Prospetto 7.1 - Il quadro concettuale per il comune: la pianificazione

| Quadro di riferimento | Strumenti Informativi |
|---|--|
| <u>Problema</u> <ul style="list-style-type: none"> ▪ Riduzione del rischio delle decisioni in ambiente mutevole ▪ Anticipazione di nuovi comportamenti ▪ Anticipazione di nuovi modelli gestionali | <u>Scenari</u> <ul style="list-style-type: none"> ▪ Comportamento degli attori ▪ Evoluzione del sistema demografico, sociale, economico ▪ Sono dati gli scenari superiori (provincia, regione, nazione) |
| <u>Schema di analisi</u> <ul style="list-style-type: none"> ▪ Attori ▪ Strategie ▪ Leve disponibili | <u>Fonti</u> <ul style="list-style-type: none"> ▪ Archivi strutturali ▪ Previsioni ▪ Interviste a testimoni privilegiati |

Prospetto 7.2 - Il quadro concettuale per il comune: la programmazione

| Quadro di riferimento | Strumenti Informativi |
|--|--|
| <u>Problema</u> <ul style="list-style-type: none"> ▪ Riprogettazione dei servizi lungo le direttrici della pianificazione | <u>Indicatori indiretti</u> <ul style="list-style-type: none"> ▪ Evoluzione storica ▪ Analogia ▪ Comportamento degli "innovatori" |
| <u>Schema di analisi</u> <ul style="list-style-type: none"> ▪ Modelli ▪ Dati ▪ Leve disponibili | <u>Fonti</u> <ul style="list-style-type: none"> ▪ Sistemi informativi interni e esterni per ricavare parametri per la programmazione dei servizi |

Prospetto 7.3 - Il quadro concettuale per il comune: la gestione

| Quadro di riferimento | Strumenti Informativi |
|---|---|
| <u>Problema</u> <ul style="list-style-type: none"> ▪ Efficienza nell'allocazione delle risorse con riferimento alle attività programmate | <u>Tecniche del controllo di gestione</u> <ul style="list-style-type: none"> ▪ Indicatori di mezzo ▪ Indicatori di risultato ▪ Indicatori di funzionamento |
| <u>Schema di analisi</u> <ul style="list-style-type: none"> ▪ Analisi dello scostamento tra programmato e realizzato | <u>Fonti</u> <ul style="list-style-type: none"> ▪ Sistemi informativi interni, dati storici ▪ Controllo di gestione |

Uno degli elementi critici è dato dalla disponibilità di fonti di dati adeguate per quanto riguarda le unità di analisi (individui e famiglie), gli ambiti informativi (economico e sociale), il dettaglio territoriale (che deve seguire la realtà dei fenomeni e non i confini amministrativi), e infine tempestività nel rilascio dei dati e nell'aggiornamento.

Il sistema ArchIMEDe ha le caratteristiche per diventare una fonte utile o indispensabile a livello locale: una vera e propria infrastruttura statistica in grado di soddisfare le esigenze informative a tutti e tre i livelli illustrati.

Il capitolo prende le mosse da questa considerazione e mostra alcuni progetti emblematici in quanto esemplificativi dell'uso di ArchIMEDe in funzione delle esigenze informative locali.

7.3 Statistica per il governo locale: dati e applicazione

Presentiamo di seguito tre esempi³ di utilizzo dell'archivio "Condizioni socio-economiche delle famiglie – ArchIMEDe" relativi ai tre livelli della pianificazione (con l'analisi della fecondità delle donne bresciane), della programmazione dei servizi (con l'analisi delle caratteristiche della invalidità civile) e infine della gestione (con l'analisi delle politiche tariffarie per alcuni servizi di base). Accanto ad elaborazioni individuali verranno presentate alcune analisi sulle famiglie.

7.3.1 Pianificazione: la fecondità delle bresciane

Nell'ambito della pianificazione in ambito sociale risulta particolarmente utile la costruzione del quadro di riferimento demografico, in modo che i decisori possano valutare l'impatto sulla domanda dei servizi della dinamica della popolazione e delle famiglie.

In questo contesto l'analisi differenziale della fecondità delle donne bresciane fornisce una informazione utile soprattutto per comprendere l'andamento della natalità, che impatta direttamente su alcuni servizi primari erogati dal Comune: si pensi anche solo alla gestione degli asili nido e delle scuole dell'infanzia.

L'analisi della fecondità è stata effettuata a partire dai dati di ArchIMEDe ricostruendo la discendenza delle donne residenti di età tra 20 e 44 anni relativamente ai figli ancora in vita e coresidenti di età tra 0 e 15 anni⁴. I dati ottenuti vanno interpretati con cautela in quanto il metodo di ricostruzione della discendenza sottoestima i valori effettivi, ad esempio il numero di figli, in particolare per le donne straniere.

La figura 7.1 rappresenta due indicatori della fecondità (numero medio di figli e età media della madre alla nascita del primo figlio) per le donne di cittadinanza italiana e delle principali comunità e aree geografiche di provenienza (definite sulla base della cittadinanza della donna). Ciò che emerge in modo immediato è la eterogeneità dei comportamenti in relazione alla provenienza delle donne, con indicatori che collocano le donne di cittadinanza italiane al limite superiore dell'età media alla nascita del primo figlio e con un numero medio di figli intermedio rispetto a quello delle altre comunità. Ad uno sguardo più approfondito emergono altri aspetti di interesse tra cui la similarità di comportamento riproduttivo per le donne dell'est Europa e, all'opposto, la dissimilarità di comportamento delle donne del subcontinente indiano e del Pakistan.

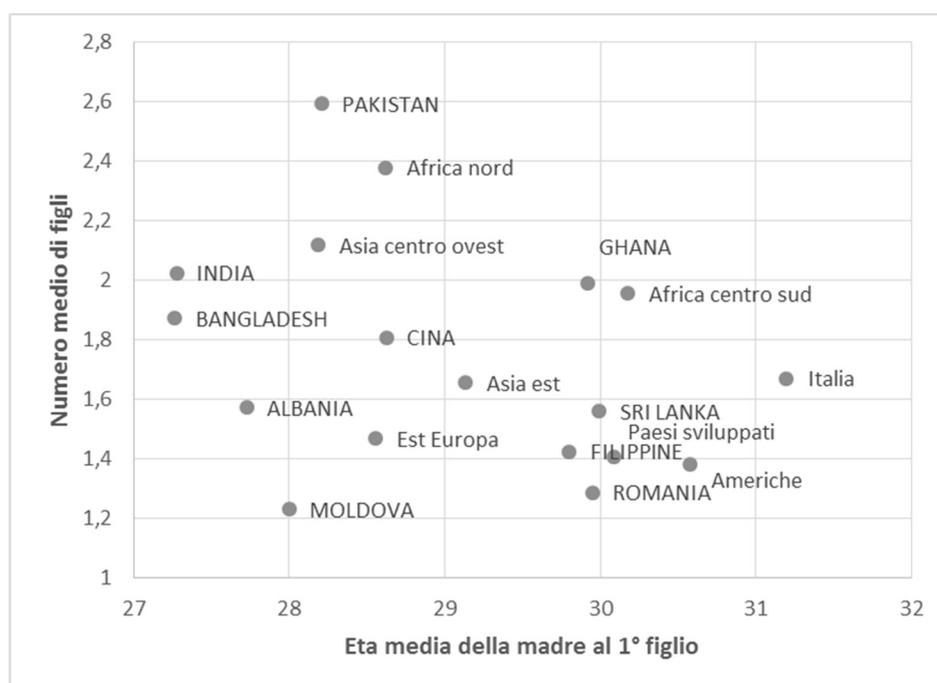
I dati mostrano vari aspetti di interesse soprattutto in una logica di valutazione prospettica della fecondità. In particolare notiamo come il livello della fecondità sperimentato dalle donne straniere residenti a Brescia sia più basso rispetto a quello rilevato nei paesi di provenienza, segno di quella tendenza delle donne migranti a rivedere al ribasso i progetti riproduttivi a fronte delle difficoltà della vita nel nuovo paese.⁵

³ Le elaborazioni sono da considerarsi sperimentali in quanto i processi di produzione di ArchIMEDe non sono ancora completamente standardizzati e normalizzati. La qualità dei dati è forse il problema maggiore, sicché alcune informazioni, ad esempio l'invalidità civile venne poi esclusa dai rilasci successivi.

⁴ La ricostruzione di storie riproduttive a partire da dati per contemporanei richiede molte cautele per varie ragioni: non tutti i figli, anche di quelle età, sono residenti con la madre; la madre a volte non è coresidente con i propri figli. Mentre sul primo punto nulla si può dire, sul secondo possiamo dire che le persone di età da 0 a 15 per le quali non è stato possibile individuare la madre, nella stessa famiglia di residenza, è del 5 per cento con differenze significative tra italiane e straniere.

⁵ Il caso della comunità ghanese, che ha comportamenti molto simili a quelli delle donne italiane, è forse quello più eclatante, e ha come spiegazione principale il fatto che quella comunità è una di quelle di più antico insediamento a Brescia.

Figura 7.1 - Distribuzione delle donne delle principali comunità cittadine per numero medio di figli e età media alla nascita del primo figlio. Comune di Brescia



7.3.2 Programmazione: l'invalidità civile

La programmazione fa riferimento alla riprogettazione dei servizi nell'arco di tempo della programmazione economica e finanziaria. In questo ambito la statistica pubblica ha come obiettivo fornire elementi per stimare la domanda potenziale dei servizi, da valutare in relazione alla domanda effettiva e alla domanda effettivamente soddisfatta dall'ente tramite i propri servizi o politiche.

L'esempio che viene presentato riguarda l'analisi dell'incidenza della disabilità a livello individuale e familiare nella prospettiva del comune come soggetto erogatore di servizi. Occorre dire che, normalmente, un comune non dispone di informazioni quantitative per stimare la domanda potenziale dei servizi per disabili, anche perché non dispone di dati sulla incidenza della disabilità nella popolazione⁶.

Una buona *proxy* dell'informazione sulla disabilità è data dalla invalidità civile che è un dato amministrativo, risultato di un processo di valutazione sanitario effettuato da amministrazioni pubbliche, e presente in ArchIMEDe, almeno nel rilascio 2012⁷.

⁶ Il comune, al più, dispone di informazioni sulla domanda effettiva, cioè sugli individui o famiglie che hanno avanzato specifiche richieste di servizi.

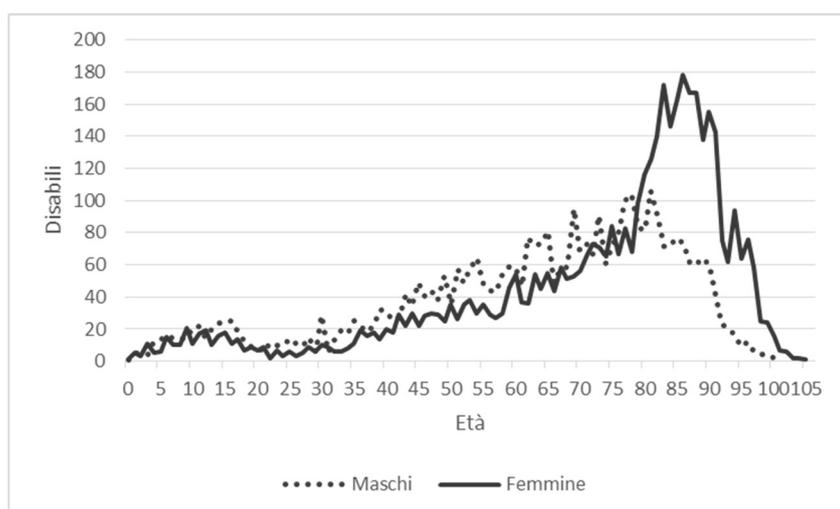
⁷ Ovviamente l'invalidità civile è una buona *proxy* della disabilità se viene riconosciuta a tutti i disabili, cioè se tutti i disabili vengono inviati alle commissioni e ad essi viene riconosciuta l'invalidità. Mentre la disabilità è uno stato, l'invalidità è un riconoscimento amministrativo necessario per poter usufruire dei benefici di legge, e richiede una domanda da parte della persona interessata, o dei familiari. Pertanto è possibile che non tutti i disabili, soprattutto quelli dalla nascita, vengano riconosciuti come invalidi perché le famiglie (per i più vari motivi, ma essenzialmente perché non ve ne è un beneficio "amministrativo") non avanzano la richiesta di valutazione alla commissione Inps. Inoltre il momento di riconoscimento della invalidità è posteriore al momento in cui si manifesta la disabilità, come è evidente nel caso delle persone anziane.

In particolare in ArchIMEDe erano disponibili informazioni sulla attribuzione di una pensione di disabilità, dove la disabilità può essere assoluta o parziale, definitiva o temporanea, e sulla presenza della indennità di accompagnamento⁸.

I dati consentono di elaborare il tasso di incidenza della invalidità civile (come *proxy* della disabilità) rispetto a due variabili rilevanti per la programmazione dei servizi, vale a dire l'età e il genere. La figura 7.2 mostra chiaramente come siano presenti due grandi cause distintive: la prima è determinata dalla nascita di persone con disabilità, che grazie agli interventi medici, sanitari e in genere assistenziali, sopravvivono nel tempo. La seconda causa, che è anche la più importante in termini quantitativi, è legata alla progressione delle condizioni invalidanti nella popolazione anziana, e in specifico nella popolazione anziana femminile.

I dati evidenziano, oltre all'aspetto quantitativo, che per altro per Brescia è in linea, per la componente giovanile, con le risultanze dei sistemi informativi dei servizi sociali, l'elemento qualitativo che rinvia a strategie e costi diversi di intervento.

Figura 7.2 - Distribuzione per età e genere degli invalidi civili. Comune di Brescia
(Valori assoluti)



Di interesse per il sistema dei servizi è l'incidenza della invalidità sulle famiglie differenziate per tipologia, dove questa non ha solo valore descrittivo, ma definisce la rete primaria (cioè parentale), di supporto alla persona invalida (Figura 7.3).

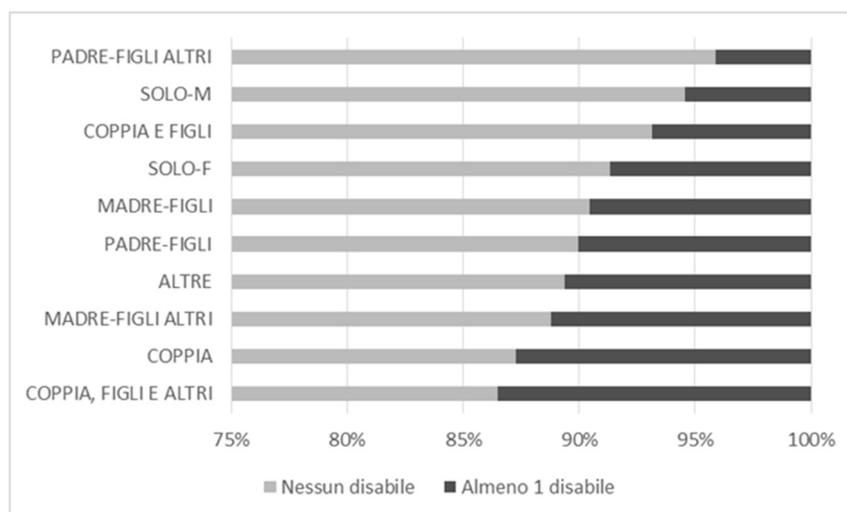
I dati, anche a questo livello di dettaglio, consentono di valutare almeno in termini quantitativi, la possibilità di tenuta della rete familiare come rete di assistenza primaria. In altri termini una situazione di equilibrio per la famiglia è rappresentata dalla presenza per ogni invalido, di almeno una persona coresidente che si possa prendere in carico il familiare bisognoso.

Seguendo questa impostazione sono preoccupanti le situazioni delle persone sole, che non hanno rete coresidente, anche per l'incidenza relativa: ben il 10 per cento delle femmine sole è costituita da persone invalide, contro il 5 per cento dei

⁸ L'attribuzione della indennità di accompagnamento permette di distinguere il livello di gravità della disabilità, in quanto individua quanti tra i disabili hanno un'invalidità totale e permanente del 100 per cento a causa della impossibilità di deambulare senza l'aiuto permanente di un accompagnatore.

maschi. Di rilievo è il dato relativo alle coppie, tipicamente anziane, tra le quali ben oltre il 10 per cento ha uno dei due componenti invalido.

Figura 7.3 - Famiglie per tipologia familiare con presenza di almeno un disabile. Comune di Brescia (Valori percentuali)



Risulterà chiaro dai dati anche sommari come per la programmazione dei servizi sia necessario differenziare in relazione non solo alle condizioni individuali, anche solo guardando alla rete familiare come rete di assistenza. La disponibilità di dati reddituali in ArchIMEDe consentirebbe di fare un ulteriore passo nell'analisi differenziale familiare per individuare le situazioni nelle quali alla debolezza della rete familiare si aggiunge la debolezza economica. È evidente come queste considerazioni abbiano a che fare con le priorità degli interventi, con i *budget* e quindi con la programmazione dei servizi.

7.3.3 Gestione: analisi dei profili tariffari per la scuola primaria

L'informazione statistica a supporto della gestione, e del controllo della stessa ha come obiettivo fornire elementi per valutare l'efficacia, in primo luogo, e l'efficienza dei servizi una volta definiti i parametri guida degli stessi. ArchIMEDe consente di disporre di informazione integrate di carattere demografico, sociale e reddituale (principalmente dai sistemi informativi dell'Agenzia delle Entrate) che sono utili per specificare i profili della domanda potenziale, da confrontare poi con la domanda effettiva o con le scelte dell'amministrazione.

L'esempio presentato riguarda l'utilizzo dei dati di ArchIMEDe per l'analisi dei profili tariffari per i servizi delle scuole primarie, frequentate dalla totalità dei residenti di età 6-10 anni.

Il comune mette a disposizione delle famiglie con figli che frequentano la scuola primaria, nell'ambito del diritto allo studio, vari servizi tra cui il servizio di ristorazione⁹.

⁹ A titolo informativo, le tariffe di tale servizio sono differenziate in relazione al reddito Isee annuo della famiglia individuando tre fasce:

- esenzione totale, per famiglie con reddito Isee fino a 17.025 euro;
- esenzione parziale, per famiglie con reddito Isee da a 17.025,01 euro a 20.220 euro;
- spesa e totale carico dell'utenza per famiglie con reddito Isee oltre a 20.220 euro.

Le tariffe del servizio sono differenziate in relazione al reddito Isee delle famiglie (che tiene solo parzialmente in considerazione la composizione qualitativa delle famiglie stesse) e non la tipologia familiare, vista anche come rete di fornitura di servizi ai componenti. Il sistema tariffario potrebbe pertanto provocare effetti distributivi e distorsivi “non voluti”, e non valutabili.

ArchIMEDe può invece fornire una base informativa per valutare quegli effetti cui si accennava elaborando la distribuzione delle famiglie per tipologia e per la fascia Isee.

La base dati originaria di ArchIMEDe non comprende l’informazione sul reddito Isee per cui abbiamo proceduto con una stima senza la componente patrimoniale.

Abbiamo calcolato una *proxy* del reddito Isee (di seguito reddito Isee*) utilizzando i dati disponibili¹⁰, che non comprendono la componente patrimoniale.

Tabella 7.1 - Distribuzione delle famiglie con bambini di età 6-10 anni per fasce di reddito annuo Isee*, secondo le fasce tariffarie comunali. Comune di Brescia (Valori assoluti e percentuali)

| Tipologia | Numero di famiglie | Reddito Isee* fino a 17.025€ | Reddito Isee* da 17.025,01€ a 20.220€ | Reddito Isee* oltre 20.220€ |
|-----------------------|--------------------|------------------------------|---------------------------------------|-----------------------------|
| Coppia e figli | 4.164 | 50,4 | 10,5 | 39,1 |
| Coppia, figli e altri | 399 | 73,4 | 8,8 | 17,8 |
| Madre e figli | 880 | 83,0 | 5,9 | 11,1 |
| Madre, figli e altri | 281 | 70,8 | 8,2 | 21,0 |
| Padre e figli | 140 | 49,3 | 10,0 | 40,7 |
| Padre, figli e altri | 416 | 60,6 | 13,0 | 26,4 |
| Altra tipologia | 180 | 86,1 | 1,7 | 12,2 |

Nota: Reddito Isee stimato senza componente patrimoniale.

Nella tabella 7.1 presentiamo il numero di famiglie con bambini di età 6-10 anni, età per la scuola primaria, distribuite nelle tre fasce tariffarie utilizzando il reddito Isee*.

La tipologia della madre sola con figli merita una particolare attenzione in relazione sia alla numerosità, si tratta di 880 famiglie con bambini di età 6-10 pari al 13,6 per cento, sia al livello del reddito che in 8 casi su 10 è inferiore alla prima fascia Isee*, ad indicare la situazione con maggiore difficoltà economica, con l’eccezione della categoria residuale. La situazione dei padri soli con figli è, invece, decisamente migliore.

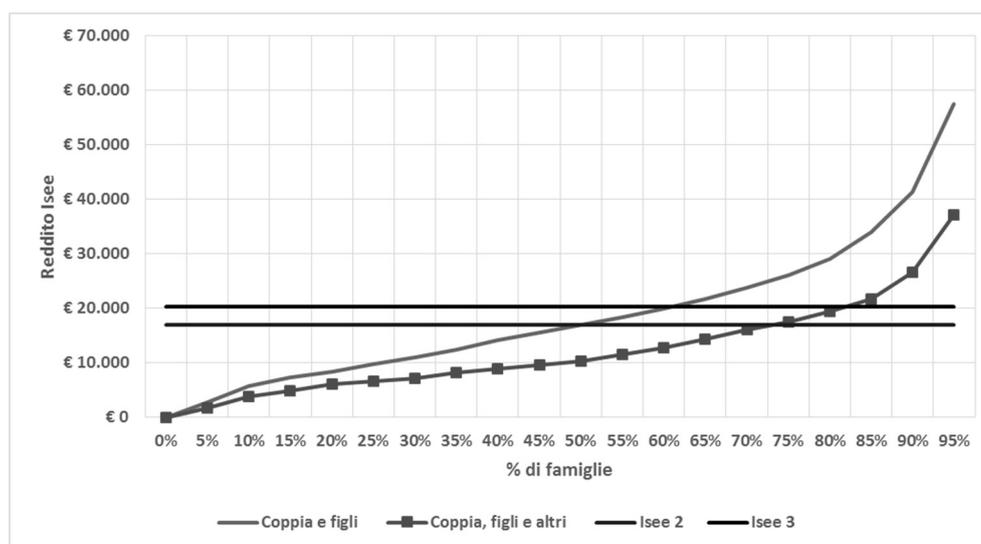
Oltre il 50 per cento delle coppie con figli (che pesano per il 64,5 per cento sul totale delle famiglie con bambini di età 6-10 anni) ha un reddito Isee* inferiore alla prima soglia, contro il 70 per cento delle coppie con figli e altri, e tra i 17mila e i 20mila euro c’è un ulteriore 10 per cento delle coppie per entrambe le situazioni. La situazione delle coppie con figli e altri è peggiore dal punto di vista reddituale rispetto alle coppie con figli. Una quota rilevante di questa tipologia è formata da famiglie composte da più nuclei, cioè dalla coppia dei genitori che coabitano con uno o più figli o figlie, e eventualmente nuora o genero, e i nipoti, tra cui il bambino in età da scuola primaria. Il fatto che la distribuzione delle famiglie di questo

¹⁰ L’indicatore Isee* viene calcolato come l’indicatore originario utilizzando una scala di equivalenza quantitativa (1=1; 2=1,57; 3=2,04; ...) a cui vengono aggiunti dei correttivi qualitativi, ad esempio, per la presenza di minori nel nucleo familiare, per la presenza di un solo genitore e per la presenza di disabili.

tipo sia più schiacciata verso redditi bassi ci indica come l'aggregazione familiare abbia una finalità di sostegno alla coppia o al soggetto più debole, cioè al più giovane.

Nella figura 7.4 per i due tipi di maggiore rilievo quantitativo (coppie con figli e coppie con figli e altri) mostriamo la distribuzione delle famiglie per decili di reddito Isee* in modo da comprendere meglio gli aspetti distributivi e forse anche distorsivi delle fasce tariffarie fissate.

Figura 7.4 - Famiglie con bambini di età 6-10 anni. Livelli di reddito Isee* per decili di famiglie. Coppie con figlie e coppie con figlie e altri. Comune di Brescia



I dati presentati evidenziano la particolarità del tipo familiare della coppia con figli e altri, che risulta avere redditi più bassi rispetto alla coppia con figli.

Ciò è spiegato dal fatto che le altre persone presenti non sono adulti, ma potrebbero essere gli stessi minori.

Ne esce un quadro di quel tipo familiare decisamente più fragile, anche economicamente, rispetto a quello che parrebbe uscire da una lettura sommaria dei dati.

7.4 Conclusioni

La sperimentazione presentata ha evidenziato potenzialità e criticità dell'utilizzo di ArchIMEDe come fonte dei dati per elaborazioni a supporto delle politiche locali. In questo paragrafo conclusivo esaminiamo tali aspetti nel dettaglio.

ArchIMEDe è un prodotto di un sistema che, grazie alla progettazione e controllo dell'intero processo produttivo, garantisce livelli elevati di qualità dei dati, e consente quindi agli utilizzatori di concentrarsi sui possibili utilizzi degli stessi.

La natura amministrativa di ArchIMEDe costituisce un aspetto che merita, invece, un approfondimento.

Il fatto che i dati derivino da sistemi amministrativi potrebbe costituire un limite in quanto è noto che le registrazioni amministrative riproducono una visione, mediata dalle specifiche regole, della realtà. Tuttavia in un ambito di supporto a processi decisionali il dato amministrativo ha una valenza propria e particolare, in

quanto la traduzione operativa delle scelte politiche non può che avvenire utilizzando informazioni amministrative. Per fare un esempio le politiche tariffarie che discriminano sul reddito, non possono che basarsi sulle informazioni amministrative sul reddito, quindi sui dati della Agenzia delle entrate.

Disporre, quindi, di una base di dati coerente con le informazioni effettivamente utilizzabili per le politiche, permette simulazioni maggiormente realistiche. Ovviamente la disponibilità di informazioni sulle situazioni reali, permetterebbe di cogliere anche eventuali aspetti distorsivi.

Un secondo aspetto di interesse è dato dalla disponibilità di informazioni relativamente a tre ambiti rilevanti per le politiche locali, vale a dire demografico, sociale e economico, e con un livello di dettaglio dei dati che non ha pari, nemmeno utilizzando i microdati censuari.

L'anonimizzazione dei dati rende invece ArchIMEDe meno facilmente integrabile con le fonti locali, soprattutto pensando di utilizzare metodi di *linkage* deterministico e del tipo uno-a-uno.¹¹ Si riduce, in tal modo, la possibilità di integrare i dati con informazioni derivanti dai sistemi informativi degli enti, che coprono aree di interesse per le amministrazioni (ad esempio i servizi sociali) e non sono disponibili nelle basi di dati nazionali. L'integrazione consentirebbe di ottenere da un unico archivio integrato sia informazioni sulla domanda potenziale sia informazioni sulla domanda effettiva, o effettivamente soddisfatta dall'ente, completando così il quadro informativo.

Infine è evidente, almeno a livello locale, come la continuità nel rilascio dei dati e la tempestività (intesa come riduzione del *lag* tra la data di riferimento e data di rilascio) siano le condizioni che permetterebbero ad ArchIMEDe di supportare efficacemente le attività degli uffici di statistica comunali e non solo.

In particolare la tempestività ha un ruolo così rilevante a livello locale, da spingere i decisori ad sacrificare addirittura la qualità del dato. Al di là di una enfasi, a volte eccessiva, sul tema, la questione per la rilevanza operativa e strategica, dovrebbe essere considerata nei processi di produzione dei dati, immaginando, ad esempio, soluzioni "statistiche", come il rilascio di archivi di dati provvisori basati su stime.

¹¹ Potrebbero essere utilizzati metodi statistici, che, tuttavia, richiedono conoscenze metodologiche e tecniche sofisticate che non sempre sono disponibili a livello locale.

CAPITOLO VIII
L'ESPERIENZA DELLA REGIONE LOMBARDIA PER LA
PREVISIONE DELLA DOMANDA DI LAVORO E LA
DEFINIZIONE DELL'OFFERTA FORMATIVA¹
di Federico Rappelli e Vincenzo Ricciari

8.1 Premessa

L'attenzione del *policy maker* è sempre concentrata sulla necessità di allocare risorse scarse in modo ottimale, specie in fasi recessive come quelle che il nostro Paese ha attraversato recentemente. Ciò si traduce nella necessità di selezionare opportunamente gli interventi da sostenere o da attivare, con un finanziamento equilibrato rispetto al fabbisogno.

In un ciclo ideale della programmazione le decisioni poggiano su informazioni di contesto e di monitoraggio (o addirittura di valutazione, nei casi più virtuosi) delle azioni implementate in precedenza. Questo è quanto avvenuto in Lombardia per quanto concerne il dimensionamento dell'offerta formativa nelle Province: ogni anno la Regione – grazie anche alle analisi effettuate dall'Istituto Éupolis Lombardia su iscritti alla Istruzione e Formazione Professionale Regionale (IeFP nel seguito) per qualifica professionale e avviamenti al lavoro a 6, 12, 18, 24 mesi dal conseguimento del titolo di studio – emette una delibera in cui tratteggia quali saranno i futuri scenari di occupazione per i formati IeFP, ovvero le probabilità di ingresso e stabilizzazione sul mercato del lavoro nonché le figure che risultano “sature”, suggerendo quindi di fatto anche la direzione nella quale orientare le scelte per il finanziamento dei percorsi formativi.

Ovviamente il panorama reale è molto meno lineare di quanto non sia quello teorico, occorre infatti tenere conto di:

- un contesto normativo in continua trasformazione;
- tempi di valutazione dilatati rispetto ai tempi dell'offerta ed erogazione dei servizi;
- oggetti di valutazione che cambiano nel tempo;
- sistemi di valutazione che non sono implementati già dall'avvio della politica.

Il tentativo che è stato fatto da Éupolis Lombardia è stato quello di usare dati rilevati a fini amministrativi (p.es. iscritti ai corsi e comunicazioni obbligatorie) per effettuare analisi utili a illustrare scenari futuri permettendo alla Regione di dare indicazioni sull'offerta di corsi IeFP. Negli anni l'uso del dato è stato affinato, sia dal punto di vista tecnico, passando dal vecchio “aggancio manuale” delle basi dati ad un processo più automatizzato sia dal punto di vista qualitativo, aumentando la profondità di analisi e le connessioni con fonti esterne, tra le quali ArchIMEDe, per

¹ Questo capitolo è debitore del lavoro presentato dagli autori alla XXXVIII Conferenza scientifica annuale AISRe su “L'uso dei dati amministrativi a supporto della programmazione delle politiche in Lombardia: l'analisi degli esiti occupazionali degli studenti della IeFP come strumento per adeguare l'offerta formativa alla domanda di lavoro.” e disponibile sul sito dell'Associazione.

arrivare sugli esiti occupazionali a un indicatore standard riconosciuto da tutti i soggetti che si occupano di mercato del lavoro in Lombardia.

8.2. Dati e metodo

I dati di fonte amministrativa utilizzati per questo studio sono di proprietà di Regione Lombardia² e vengono rilevati per le necessità legate alla programmazione e alla gestione delle attività; per un loro utilizzo per finalità statistiche, devono essere prima adeguatamente trattati.

Più nel dettaglio, i dati cui facciamo riferimento sono quelli relativi a:

- iscritti, qualificati e diplomati nei percorsi di Istruzione e Formazione Professionale di livello regionale, con dettaglio di età, genere, provincia di residenza, qualifica professionale, percorso formativo (promossi/bocciati per ogni passaggio annualità) ed esito finale (voto di uscita);
- percorsi offerti, per qualifica professionale, sede di corso, comprese le informazioni anagrafiche del soggetto erogatore;
- comunicazioni obbligatorie (avviamenti, avviati, cessazioni, cessati, proroghe, trasformazioni) per età, sesso, qualifica professionale (CP2011, Istat), titolo di studio, tipologia oraria (tempo pieno/tempo parziale), fattispecie contrattuale (tirocinio, apprendistato, tempo determinato, tempo indeterminato, lavoro occasionale e accessorio, con le relative declinazioni ulteriori), sede del datore di lavoro a livello, domicilio del lavoratore a livello comunale, codice della classificazione delle attività economiche (Ateco) del datore di lavoro;
- politiche erogate da Regione: Garanzia Giovani, Dote Unica Lavoro, Dote scuola, con informazioni sulla presa in carico (data, durata, servizio usufruito) e sull'utente (età, sesso, titolo di studio, stato di partenza, Neet, disoccupato, inoccupato, fascia di vulnerabilità).

Affinché ci sia comparabilità tra anni e figura professionale, è necessario che le qualifiche del repertorio regionale vengano ricondotte alla classificazione Istat delle professioni (CP2011, Istat a 3 e 5 *digit*) secondo gli abbinamenti illustrati in tabella 8.1, a partire dalle competenze in essere richieste e considerando le modifiche apportate al repertorio regionale.

² In particolare, i dati sono raccolti dalla DG Istruzione Formazione e Lavoro nell'ambito delle attività legate alla erogazione delle Doti. Si ringraziano, qui nuovamente, la DG e i Funzionari regionali per la collaborazione.

Tabella 8.1 - Abbinamento figure del repertorio delle qualifiche IeFP con le qualifiche professionali CP2011 Istat (3 digit)

| <i>Figura regionale</i> | <i>Qualifiche CP2011 Istat - 3 digit</i> |
|--|--|
| Operatore amministrativo segretariale | <ul style="list-style-type: none"> ▪ Impiegati addetti alla segreteria e agli affari generali ▪ Impiegati addetti all'archiviazione e conservazione della documentazione ▪ Personale non qualificato di ufficio ▪ Impiegati addetti alla gestione amministrativa della logistica ▪ Impiegati addetti alla gestione economica, contabile e finanziaria ▪ Impiegati addetti al controllo di documenti e allo smistamento e recapito della posta |
| Operatore ai servizi di vendita | <ul style="list-style-type: none"> ▪ Addetti alle vendite ▪ Altre professioni qualificate nelle attività commerciali |
| Operatore della ristorazione | <ul style="list-style-type: none"> ▪ Esercenti ed addetti nelle attività di ristorazione |
| Operatore del benessere | <ul style="list-style-type: none"> ▪ Operatori della cura estetica |
| Area meccanica | <ul style="list-style-type: none"> ▪ Fonditori, saldatori, lattonieri, calderai, montatori di carpenteria metallica e professioni assimilate ▪ Operatori di catene di montaggio automatizzate e di robot industriali ▪ Operai addetti a macchine automatiche e semiautomatiche per lavorazioni metalliche e per prodotti minerali ▪ Fabbri ferrai costruttori di utensili ed assimilati ▪ Meccanici artigianali, montatori, riparatori e manutentori di macchine fisse e mobili (esclusi gli addetti alle linee di montaggio industriale) |
| Operatore elettrico e operatore elettronico | <ul style="list-style-type: none"> ▪ Artigiani e operai specializzati dell'installazione e della manutenzione di attrezzature elettriche ed elettroniche |
| Operatore della trasformazione agroalimentare | <ul style="list-style-type: none"> ▪ Artigiani ed operai specializzati delle lavorazioni alimentari ▪ Operai addetti a macchinari fissi per l'industria alimentare |

8.3 Risultati

8.3.1 Il dimensionamento dell'offerta formativa nelle province lombarde

Al fine di rendere il mercato del lavoro più dinamico e inclusivo e ridurre il *mismatch* tra domanda e offerta di lavoro, ovvero adeguare la formazione (ad ogni livello e per ogni competenza) ai fabbisogni del tessuto economico del territorio e dunque accrescere le possibilità di inserimento lavorativo (per qualità, rapidità di inserimento e stabilità), Regione Lombardia ha implementato negli ultimi anni un sistema di monitoraggio utile a rivedere di anno in anno l'offerta formativa professionale erogata nei territori.

In accordo con la Regione, Éupolis Lombardia ha prodotto documenti analitici di supporto, analizzando per ogni provincia le qualifiche professionali di iscritti e formati (III e IV anno) nel sistema IeFP, gli esiti occupazionali dei formati (quanti lavorano, in quali settori, con quali contratti) e la coerenza tra percorso formativo e lavorativo; la lettura congiunta di questi fattori offre alle province una maggiore conoscenza del mercato del lavoro locale, ovvero consente di modulare l'offerta formativa tenendo conto di fattori cruciali per la formazione professionale.

Uno dei fattori di interesse risiede nel fatto che nel corso delle diverse edizioni è emersa la necessità di approfondire il dettaglio delle analisi e allargarne lo spettro per dare una "spiegazione" delle scelte formative dei ragazzi e delle relative ricadute occupazionali delle figure professionali, con l'obiettivo di anticipare le necessità dei territori e delle imprese, ovvero prevedere quali possano essere le figure professionali maggiormente ricercate in futuro in generale (nel breve/medio periodo) e

quali invece risultino “sature” (sia nella scuola che nel mercato) anche in base alle specificità territoriali e del sistema economico locale.

Per fare ciò abbiamo sviluppato un modello³ che, partendo dalle informazioni contenute nelle comunicazioni obbligatorie (avviamenti), stima il numero di avviamenti per figura professionale, mantenendo tuttavia l’obiettivo di usare queste stime per poi dare indicazioni alle province su quanti posti assegnare e quali corsi attivare per figura professionale. Il lavoro è stato strutturato nel seguente modo:

- Sistematizzazione dei dati sugli avviamenti: per singola figura professionale, qualifica CP2011 Istat a 3 digit, con dettaglio per fascia di età (17-22 anni).
- Sistematizzazione dei dati IeFP: iscritti (III e IV anno), qualificati e diplomati per figura professionale e provincia, promossi (successo formativo), ingressi sul mercato del lavoro, esiti occupazionali (6, 12, 18, 24 mesi), compreso raccordo tra le figure che risultano essere diverse tra un anno e l’altro (per modifiche nel repertorio regionale) e tra diplomati e qualificati (operatore vs tecnico).
- Raccordo delle figure professionali della IeFP con le qualifiche professionali Istat: al fine di individuare il bacino di riferimento nel mercato del lavoro per tutti i formati.
- Analisi mensile e trimestrale delle variazioni tendenziali e congiunturali degli avviamenti e individuazione della quota di mercato per figura (analisi della domanda di mercato).
- Analisi annuale delle variazioni e delle quote per gli iscritti IeFP (analisi dell’offerta potenziale).
- Analisi delle variabili esogene (p.es., la stima dell’elasticità dell’occupazione al PIL).
- Analisi di trend e stagionalità delle serie, medie mobili (a 3, 5, 7 periodi) e smussate ($0,75*0,25$; $0,25*0,75$, $0,5*0,5$): definizione dello scostamento accettabile (minimo e massimo), media non pesata per figura e totale avviamenti e media pesata trimestrale (i trimestri più recenti sono più predittivi e quindi pesano di più, 0,85 vs 1,15).
- Definizione dei vincoli di previsione: il trimestre di partenza è l’ultimo con valori noti, definizione dell’intervallo di previsione valida, le previsioni devono essere in linea con il trend e la stagionalità e l’andamento stimato per figura professionale.

Una volta stimato il valore assoluto (alto, medio, basso secondo i pesi del modello) degli avviamenti per i primi tre trimestri del 2017 è stata riproporzionata la quota degli avviamenti per figura professionale sul valore stimato degli iscritti al nuovo anno formativo della IeFP. Come si può notare la differenza tra la distribuzione media (che dipende dalla media regionale) e la distribuzione prevista (che dipende dagli andamenti delle figure 8.1 e 8.2) risiede fondamentalmente nel fatto che la seconda enfatizza gli andamenti dei trimestri recenti, facendo pesare di più le figure professionali che hanno un andamento crescente nel tempo, ovvero area meccanica e ristorazione a discapito dell’area dei servizi di vendita; evidenziamo tuttavia l’aumento del tutto insufficiente dell’area amministrativa e segretariale pur in presenza di una forte domanda di lavoro e – al contrario – un aumento ancora cospicuo dell’area del benessere che invece presenta una domanda di lavoro ormai satura.

³ Rappelli e Ricciari (2017).

Figura 8.1 - Avviamenti totali delle figure considerate e previsioni. Lombardia, 2017, trimestri I, II, III

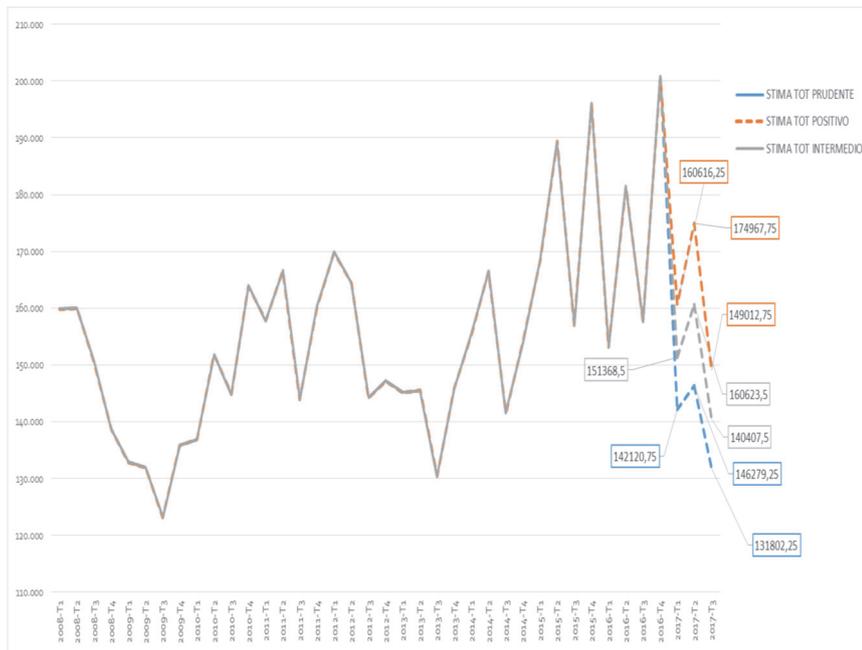
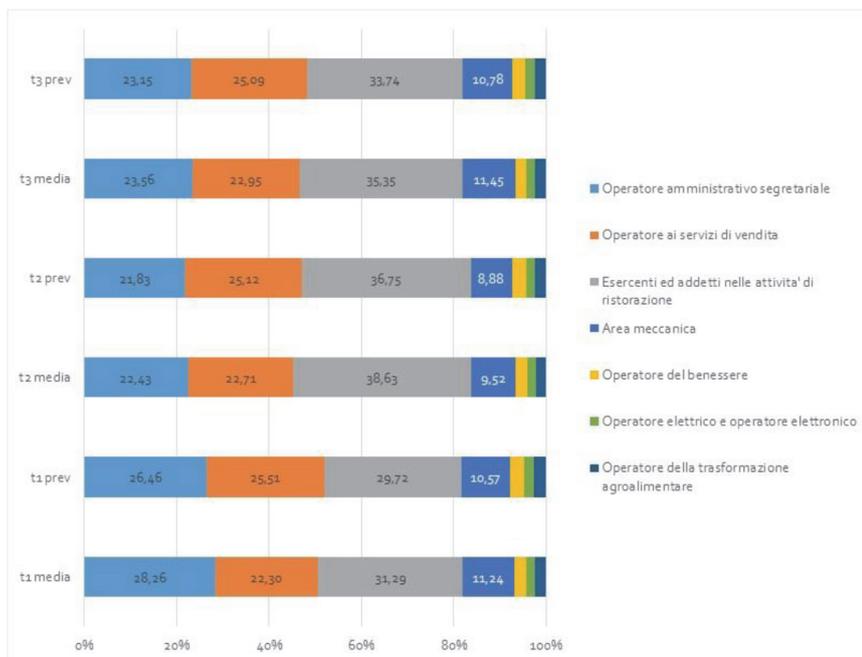


Figura 8.2 - Quote percentuali degli avviamenti totali delle figure professionali considerate, , media per tipo di trimestre e previsioni. Lombardia, 2017, trimestri I, II, III



8.3.2 Migliorare le stime associando altri dati amministrativi

Una ulteriore analisi, maggiormente articolata, è stata possibile proprio attraverso la definizione di una base informativa che è stata costruita per tracciare tutti gli eventi di istruzione, formazione e lavoro registrati nelle fonti amministrative nazionali e regionali, in modo da ricostruire la storia di ciascuna unità a partire dal suo status iniziale.

La Regione ha condiviso i dati amministrativi sulla formazione professionale e sulle comunicazioni obbligatorie (apprendistati e tirocini extra-curricolari), Istat ha provveduto all'integrazione con le basi dati provenienti da fonti amministrative nazionali (p.es. Miur e Inps – residenza, istruzione, lavoro, reddito), sotto la sperimentazione del progetto ArchIMEDe in particolar modo rispetto a *Precarietà lavorativa*⁴ e *Popolazione insistente* (Cfr. Capitolo 1).

L'oggetto di analisi in questo caso è stato la sottopopolazione dei giovani senza segnali amministrativi di istruzione, formazione e lavoro per analizzarne le principali caratteristiche socio-demografiche (sesso, età, cittadinanza, titolo di studio) e la durata della permanenza nella condizione di "senza segnali", anche ad un livello territoriale disaggregato (comune di residenza)⁵.

Il mese di ottobre è stato selezionato come periodo di riferimento per l'individuazione della popolazione presente, sia perché inizia l'anno scolastico sia perché non è soggetto a stagionalità per quanto concerne il mercato del lavoro. Le unità presenti sono ricavate da tre popolazioni: la popolazione residente in Italia, gli iscritti a corsi di istruzione e formazione ed i lavoratori. È stata dunque individuata la popolazione di riferimento finale: i giovani di età compresa fra i 15 ed i 29 anni presenti nelle regioni oggetto di analisi (con residenza o segnali di studio-lavoro in regione).

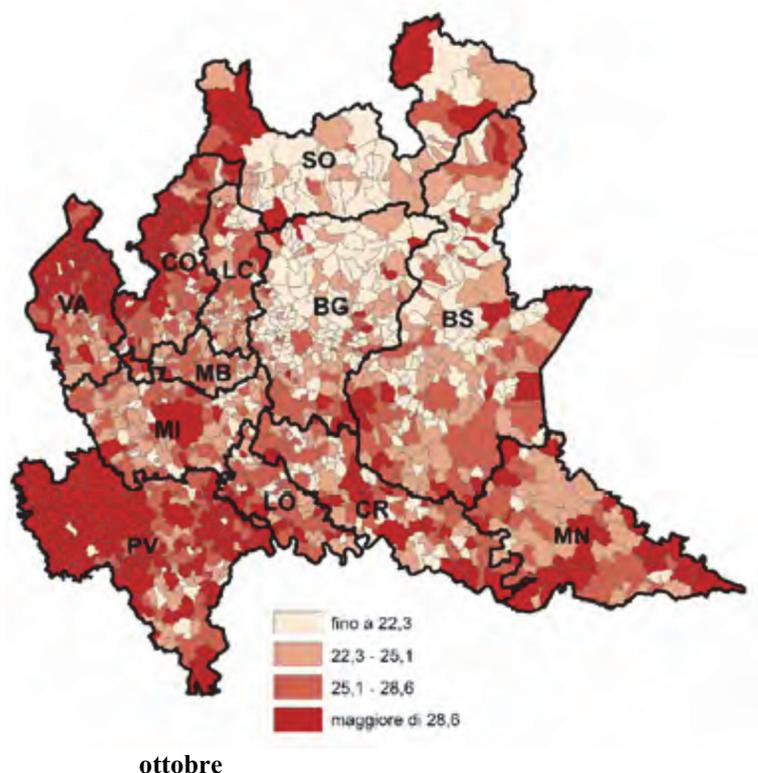
L'integrazione a livello micro di più fonti amministrative permette di avere informazioni di dettaglio sui percorsi formativi e lavorativi degli individui, consentendo di fare analisi che in precedenza non potevano essere effettuate, per esempio sapere chi sono e dove sono le persone che in un dato momento non hanno un'occupazione e non sono presenti nel sistema educativo.

Ciò che emerge sugli individui nella fascia di età 15-29 anni che non hanno segnali di formazione o lavoro (Neet) è così sintetizzabile:

- dal 2011 al 2013 la quota di Neet è rimasta sostanzialmente stabile;
- fino ai 19-20 anni non si registrano differenze di genere, dai 21 anni in su il divario tra uomini e donne aumenta a svantaggio delle donne, in modo particolare per gli stranieri (che già presentano valori superiori agli italiani);
- si registra una generale permanenza nel medesimo stato a distanza di due anni;
- all'aumentare del titolo di studio aumenta la quota dei "senza segnali" che entrano nel mercato del lavoro gli anni successivi;
- più ci si allontana dall'uscita dal sistema educativo più diminuisce la quota degli individui "senza segnali" e, in maniera complementare, aumenta la quota di individui che lavorano a tempo pieno (come conferma l'andamento dell'intensità di lavoro).

⁴ Garofalo (2014).

⁵ Filippini, Laghi e Ricciari (2017).

Figura 8.3 - Incidenza percentuale dei "senza segnali" per comune. Lombardia, 2013,

Note: Questi risultati meritano ulteriori approfondimenti anche in relazione alla tematica del lavoro oltre confine (Cfr. Capitolo 3 e Appendice 2).

8.4 Conclusioni

Le informazioni di provenienza amministrativa sono da anni ormai al centro della revisione del processo statistico, consentendo così di passare in poco tempo dal paradigma del campione all'accesso totale all'universo; se prima era importante definire fin da principio la strutturazione del campione per minimizzare gli errori di riporto all'universo e le variabili rilevate non potevano essere riutilizzate ad altri scopi (di approfondimento o di aggancio con altre rilevazioni), oggi la presenza di numerose informazioni raccolte per specifici fini (amministrativi nel caso della Regione) se adeguatamente trattate (*data cleaning*, *data quality*, verifica dell'attendibilità) e trasformate per usi statistici possono divenire fonte, robusta, di analisi e monitoraggio, a maggior ragione se agganciate ad altre fonti informative (anche se costruite a scopi differenti o riguardanti temi non direttamente connessi con l'oggetto di indagine).

Regione Lombardia ha da tempo sviluppato un sistema di monitoraggio dei dati elementari degli individui che accedono alle politiche regionali; l'Istituto utilizza queste informazioni per supportare il decisore pubblico, nello specifico grazie all'analisi degli esiti occupazionali, sia per misurare l'andamento del mercato del lavoro (comunicazioni obbligatorie) che per stimare gli effetti delle politiche (quanti lavorano dopo essere stati "aiutati" con i servizi offerti e disegnati su misura?) o per adeguare il disegno dell'offerta formativa (vedi *supra*).

In presenza di almeno una chiave di collegamento (quale il CF per esempio) i dati elementari di una amministrazione acquisiscono forza se connessi con le informazioni di altre amministrazioni (p.es. Istat, Miur, Inps, Agenzia delle Entrate); e questo è quanto è stato fatto con la sperimentazione di ArchIMEDe; un approfondimento ulteriore sull'intensità di lavoro è teso a confrontare le differenze non solo per territorio, età e cittadinanza ma anche rispetto alle qualifiche professionali (quali garantiscono maggiore facilità e rapidità di ingresso? Quali comportano maggiore stabilità? Sono le medesime ad avere successo alcune emergono alla distanza?).

Ulteriori potenzialità sarebbero sviluppabili: pensiamo ad esempio agli effetti delle origini familiari (occupazione e titolo di studio dei genitori) sulle scelte formative dei figli, uno dei nuovi filoni in corso di sviluppo, in cui sarebbe possibile confrontare le distribuzioni di frequenza di due universi per stimare le differenze territoriali (lombarde) o il successo scolastico per percorso formativo (IeFP regionale vs gli altri percorsi); così come gli effetti della provenienza dello studente (da istruzione secondaria di I livello o da altro percorso, studenti ripetenti o meno).

Infine, i dati amministrativi possono acquisire valore per produrre in tempo reale (o comunque inferiore, e di molto, a quelli di una rilevazione statistica) informazioni supplementari a quelle per cui sono raccolti, come per esempio abbiamo fatto in Regione usando i dati delle comunicazioni obbligatorie per costruire una nuova geometria dei Sistemi Locali del Lavoro (Fossi e Ricciari, 2015), normalmente modificati con cadenza censuaria, consentendo dunque al decisore di adeguare le politiche in tempi relativamente brevi.

INTRODUZIONE ALLA PARTE QUARTA
APPLICAZIONI SPERIMENTALI SU DATI AMMINISTRATIVI
di Marianna Filandri

Gli approcci allo studio della povertà sono numerosi. I più diffusi sono quelli che si focalizzano su un singolo aspetto e forniscono informazioni sintetiche sulla presenza o assenza di povertà. A questo proposito sono ben note le statistiche ufficiali Istat che vengono spesso riprese e commentate nel dibattito pubblico. La dimensione considerata è quella monetaria e l'indicatore di povertà è dicotomico: poveri/non poveri. I vantaggi di questo approccio sono diversi: l'informazione sintetica, la facilità di comprensione, la possibilità di fare confronti tra contesti diversi, il considerare i cambiamenti nel tempo. D'altro canto, ci sono anche svantaggi. Definire lo spazio valutativo della povertà identificando solo una dimensione, significa non tenere conto della complessità del fenomeno. Inoltre la misura della condizione di svantaggio può essere articolata in gradi di severità, identificando situazioni di grave o lieve deprivazione. Tenere conto di più dimensioni consente anche di rilevare le situazioni di disagio cumulativo in cui vivono alcune famiglie. Ad esempio un nucleo a basso reddito i cui membri hanno un livello di istruzione elementare e che vive in una condizione di sovraffollamento o in una abitazione non adeguata soffre contemporaneamente di povertà monetaria, culturale e abitativa. Situazioni esistenziali disagiate e insoddisfacenti richiamano l'idea che la deprivazione sia costituita congiuntamente da più aspetti – appunto quelli materiali, psicologici, culturali, relazionali e sociali. In altre parole la povertà è un fenomeno multidimensionale che configura per gli individui condizioni di esistenza difficili e spesso isolate dalle usuali reti dei rapporti culturali, sociali ed economici sui quali si fonda e nei quali consiste la vita associata (Lucchini *et al.*, 2007; Autigna e Filandri, 2015). Ma non solo. Pur focalizzandosi su un aspetto specifico, quello del reddito, e considerando il livello di povertà economica, le situazioni che si possono presentare sono difficilmente riconducibili a un'unica dimensione. Il semplice incrocio della dimensione individuale con quella familiare mostra una realtà sociale in cui sono presenti situazioni di deprivazione complessa che si inseriscono solitamente a metà tra i due poli, ossia tra la condizione di individui e famiglie che stanno sicuramente bene e quella di coloro che stanno sicuramente male. L'analisi di questi dati è diventata imprescindibile nello studio della povertà, considerando la natura del fenomeno. Come anche i contributi di questa parte del volume mostrano – attraverso l'analisi di dati amministrativi particolarmente ricchi – è dunque auspicabile considerare sempre più altre dimensioni accanto a quella monetaria, come ad esempio quella del lavoro, quella della salute e della casa. In realtà, anche la mancanza di istruzione può essere considerata, di per sé, un indicatore di povertà.

Gli studi che hanno indagato la relazione tra titolo di studio e rischio di caduta in povertà mostrano molto bene l'importanza di contestualizzare questa relazione nello spazio e nel tempo. L'area di residenza ha un effetto di interazione sulla deprivazione: in alcune aree avere un titolo di studio basso aumenta notevolmente la probabilità di cadere in una situazione di povertà economica. Questo dato, già di per sé meritevole di attenzione, è ulteriormente aggravato dalla facilità con la quale lo svantaggio si trasmette da una generazione all'altra. In letteratura è ben noto il cosiddetto effetto San Matteo (Merton, 1988) secondo cui coloro che hanno di più sono anche coloro che avranno di più. In altre parole si tratta del meccanismo per il quale le disuguaglianze si riproducono nel corso della vita e tra una generazione e l'altra. Questo significa non solo che coloro che pos-

seggono un titolo di studio basso hanno maggiori probabilità di trovarsi in povertà, ma anche che i loro figli hanno minori probabilità di acquisire alti titoli di studio e, corrono un elevato rischio di abbandono degli studi dopo la fine dell'obbligo scolastico (Struffolino e Filandri, 2015). Il contributo di Carulli, Sicuro e Tucci sulle disparità di genere nel mercato del lavoro conferma l'importanza del titolo di studio come elemento protettivo nel divario occupazionale e di reddito tra uomini e donne. La povertà può allora essere vista come una sindrome di carenze che si rinforzano l'un l'altra e presentano tratti di stabilità nel tempo, definendo un modello di deprivazione (Morlicchio, 2012). Questo modello proposto da Townsend (1987) tiene conto simultaneamente sia dell'insieme di deprivazioni materiali (alimentazione, alloggio, condizione lavorativa) sia di quello di deprivazioni sociali, ossia la mancanza di istruzione o relazioni sociali. Emergono così situazioni di radicale disagio: individui ai quali mancano risorse basilari: vivono in ristrettezze finanziarie, hanno notevoli difficoltà nell'arrivare alla fine del mese, non sono in grado di mantenere una alimentazione equilibrata, non si possono permettere di avere ospiti a cena, di comprarsi vestiti nuovi, di acquistare elettrodomestici, a maggior ragione di cambiare mobili e di andare in vacanza per almeno una settimana all'anno. Inoltre vivono in case male illuminate, umide, non adeguatamente riscaldate e, spesso, in situazioni di sovraffollamento. Frequentemente, poi, queste abitazioni sono ubicate in quartieri inquinati e afflitti da criminalità diffusa. E anche senza questi due ultimi elementi di disagio, le persone delle quali stiamo discutendo si trovano spesso in cattiva salute, soffrono di qualche malattia cronica e manifestano difficoltà a svolgere autonomamente gli atti elementari della vita quotidiana (Lucchini *et al.*, 2007). La cumolazione degli svantaggi si evince bene dalle analisi presentate nei capitoli 8 e 9, rispettivamente di Succi, Vannoni e Vitalini e di Casacci. Nel primo contributo viene proposto un indice di vulnerabilità socio-economica attraverso il quale sono identificate, con un livello di dettaglio piuttosto fine, aree di bassa o elevata fragilità. Il secondo contributo presenta anch'esso delle cartografie di aree di svantaggio che sono state messe in relazione con tipologie familiari che presentano una vulnerabilità economica maggiore rispetto ad altre. Sempre su questa linea il capitolo di Della Queva, Nicosia e Stoppiello fa un passo ulteriore e propone un'analisi sperimentale della presenza e la diffusione territoriale di alcune tipologie di istituzioni non profit, alla luce di alcune forme di vulnerabilità familiare della popolazione residente in Lombardia. Ma l'introduzione di altre dimensioni porta l'attenzione anche su nuovi scenari e la complessità dello studio della deprivazione si dispiega chiaramente. La mancanza di occupazione, ad esempio, rappresenta in generale uno dei fattori che con più forza si associa alla maggiore incidenza della povertà, ma essere occupati non è una condizione di per sé sufficiente a proteggere dalla povertà, come mostrano chiaramente Cariello e Carbonara nel capitolo sui *working poor* e Ballabio e colleghi relativamente ai *Quasi-Neet* della Lombardia. Non solo, sebbene in quest'ultimo lavoro emerga che la condizione di disoccupazione o di inattività non implichi l'assenza totale di reddito, specifici sottogruppi presentano elevati livelli di vulnerabilità in relazione non solo all'esclusione dal sistema formazione-lavoro ma anche rispetto all'assenza di una rete di protezione socio-parentale (p.es. basso reddito del nucleo familiare di origine).

In conclusione, proprio come le ricerche presentate in questa parte del volume, lo studio della povertà non dovrebbe limitarsi alla sola dimensione del reddito e dei consumi: adeguata attenzione dovrebbe essere prestata alle caratteristiche e ai fattori di svantaggio delle famiglie e delle persone identificate come povere, ma anche di quelle vulnerabili. E particolare attenzione va prestata all'interazione tra le dimensioni di svantaggio materiale e sociale. Di questo potranno beneficiare il disegno e l'attuazione di efficaci politiche di contrasto alla povertà, il cui obiettivo deve quindi essere quello di intervenire ad ampio spettro.

CAPITOLO IX

WORKING POOR NEL NORD-OVEST

di Salvatore Cariello e Monica Carbonara

9.1. Introduzione

Con l'acuirsi della crisi economica e con le politiche neoliberiste messe in atto per fronteggiarla, il *working poor*, il lavoratore povero, per il quale il lavoro non è più “una buona assicurazione contro la povertà” ha assunto sempre maggiore rilevanza sia nel dibattito politico-sindacale che nella ricerca socio-economica.

Le ricerche sui *working poor* in Italia hanno utilizzato, per la gran parte, i risultati di indagini campionarie, con i conseguenti limiti di significatività dei dati quando il livello di dettaglio, tematico o territoriale, si fa molto fine. Questo lavoro utilizza una fonte statistica dell'Istat recentemente messa a disposizione e ancora poco esplorata, i *data set* del Progetto ArchIMEDe¹, per proporre una analisi del fenomeno del *working poor* nei comuni delle regioni del Nord-Ovest con l'obiettivo di fornire elementi utili a individuare aree particolarmente esposte al fenomeno e analizzarne le principali caratteristiche.

In linea con le definizioni utilizzate a livello europeo, la soglia di povertà è stata fissata al 60 per cento del reddito mediano. Occorre, tuttavia, precisare che i microdati del progetto ArchIMEDe forniscono il reddito al lordo delle imposte e che, allo stato, mancano del tutto i redditi dei frontalieri con la Svizzera², pertanto l'attribuzione dello status di povertà prescinde dagli effetti redistributivi della tassazione e dall'apporto dei redditi derivanti dal lavoro transfrontaliero. Inoltre, poiché i microdati disponibili forniscono distintamente l'ammontare del reddito da lavoro dipendente e non quello da lavoro autonomo – perché non scorporabile da altre tipologie di reddito – l'analisi sarà condotta solo in riferimento ai lavoratori dipendenti.

L'analisi, effettuata sui *data set* ArchIMEDe 2014, adotta, quindi, le seguenti definizioni (Cfr. Capitolo 2):

- famiglia: insieme di persone legate da vincoli di matrimonio, parentela, affinità, adozione, tutela o affettivi coabitanti ed iscritte nelle anagrafi comunali;
- famiglia povera: famiglia con reddito familiare lordo equivalente inferiore al 60 per cento della mediana del reddito lordo equivalente delle famiglie presenti nella base dati di ArchIMEDe;

¹ Il Progetto ArchIMEDe ha reso disponibili collezioni di microdati, di fonte amministrativa, relativi all'universo degli individui e delle famiglie residenti nei comuni italiani. Per ogni individuo iscritto in anagrafe, sono presenti informazioni sulle caratteristiche socio-demografiche (sesso, età, titolo di studio, cittadinanza, relazione di parentela) sulla condizione lavorativa, sull'ammontare del reddito lordo e di quello da lavoro dipendente. Inoltre, ricostruiscono a partire dai dati individuali le informazioni relative alle famiglie (tipologia di famiglia, numero di componenti, reddito lordo, reddito lordo equivalente, ecc.).

² In base all'accordo Svizzera-Italia del 3 ottobre 1974, i salari, gli stipendi e gli altri elementi facenti parte della remunerazione che un lavoratore frontaliero riceve in corrispettivo di una attività dipendente sono imponibili soltanto nello Stato in cui tale attività è svolta. La qualificazione di “frontaliero” svizzero è riconosciuta ai lavoratori residenti in un comune italiano il cui territorio è compreso, in tutto in parte, nella fascia di 20 Km dal confine (Cfr. Appendice 2).

- lavoratore dipendente: individuo, iscritto in anagrafe, che nella base dati risulta avere:
 - una occupazione alle dipendenze nel mese di ottobre;
 - percepito redditi da lavoro dipendente;
 - segnali di attività lavorativa per almeno sei mesi nell'anno.
- lavoratore dipendente povero (*working poor*): lavoratore dipendente, come definito al punto precedente, con reddito da lavoro dipendente inferiore al 60 per cento della mediana dei redditi da lavoro dipendente presenti nella base dati di ArchIMEDe.

9.2 Famiglie, individui e lavoratori poveri nei comuni del Nord-Ovest

Il ricorso a dati di fonte amministrativa consente di avere, con continuità, dati con livelli di dettaglio territoriale che solo le rilevazioni statistiche totali riescono a garantire. Le differenze, metodologiche e classificatorie, sono tali da non consentire un confronto diretto con i dati tra fonti amministrative e fonti statistiche³, tuttavia i dati di ArchIMEDe possono essere utilizzati per il calcolo di misure *proxy* a livello comunale di indicatori tipicamente calcolati da dati di indagine con granularità territoriale meno fine.

9.2.1 Universo di riferimento

La popolazione *target* dell'analisi è rappresentata dai residenti nei comuni del Nord-Ovest che nella base dati ArchIMEDe risultano avere una occupazione alle dipendenze nel mese di ottobre 2014, hanno percepito reddito da lavoro dipendente e presentano segnali di lavoro per almeno sei mesi. Si tratta di 4,5 milioni di lavoratori, pari al 71,4 per cento degli occupati nel mese di ottobre 2014. Poiché si ritiene opportuno esaminare il fenomeno nel suo complesso, l'analisi qui proposta terrà conto sia della posizione dei singoli lavoratori sul mercato del lavoro sia delle caratteristiche delle famiglie in cui essi vivono.

L'incrocio fra la dimensione individuale e quella familiare porta a suddividere l'universo dei lavoratori dipendenti in quattro gruppi, secondo lo schema della figura 9.1⁴. Il primo gruppo (quadrante I) include i lavoratori dipendenti che, per livello di reddito individuale e familiare, possono essere considerati fuori dall'area del *working poor*; nel secondo si collocano quei lavoratori con redditi da lavoro dipendente superiori al 60 per cento del reddito mediano lordo da lavoro dipendente, che vivono in famiglie povere; il terzo raggruppa gli individui "poveri" dal punto di vista del reddito individuale, ma che fanno parte di famiglie con reddito lordo equivalente superiore alla soglia di povertà familiare; l'ultimo gruppo, infine, include i lavoratori dipendenti poveri che vivono in famiglie povere.

9.2.2 Il *working poor* nelle regioni del Nord-Ovest

Come si evince dalla figura 9.1, quattro lavoratori dipendenti su cinque si collocano fuori dell'area di *working poor*. Sono circa 900mila, invece, quelli che si posizionano nell'area del *working poor*. I lavoratori poveri che vivono in famiglie

³ Oltre alla diversità delle definizioni adottate (al riguardo vedi la nota metodologica consultabile all'indirizzo: <http://www.istat.it/it/archivio/190365>), occorre tener presente che, per l'anno 2014, nella base dati di ArchIMEDe sono presenti i dati relativi a 7,1 milioni di famiglie e 15,8 milioni di individui residenti in 2.983 dei 3.046 comuni del Nord-Ovest e che nei 63 comuni mancanti risiedono 302 mila individui, pari all'1,8 per cento della popolazione residente.

⁴ Lo schema di classificazione è ripreso da Filandri e Struffolino (2012).

non povere sono il gruppo più numeroso (11,7 per cento), segue il gruppo dei lavoratori che si collocano al di sotto della soglia di povertà sia in termini di reddito individuale che familiare (5,8 per cento) e una piccola quota (2,2 per cento) di occupati non poveri che vivono in famiglie povere.

Figura 9.1 - Relazioni povertà lavoratore dipendente – povertà familiare e delimitazione dell'area *working poor*

| | | Lavoratore dipendente | | |
|----------|------------|---|---|---|
| | | Non povero | Povero | |
| Famiglia | Non povera | I 80,2% | III – Lavoratore povero in famiglia non povera 11,7% | <input type="checkbox"/> No <i>Working poor</i> |
| | Povera | II – Lavoratore non povero in famiglia povera 2,2% | IV – Lavoratore povero in famiglia povera 5,8% | <input type="checkbox"/> <i>Working poor</i> |

Attraverso un modello di regressione logistica, costruito con la procedura *backward elimination* ed escludendo l'intercetta nel modello di regressione, sono state individuate le principali determinanti della povertà. In linea con la letteratura e le ricerche sul tema, l'analisi ha evidenziato l'incidenza di fattori individuali – il sesso, l'età, il livello di istruzione, la stabilità dell'occupazione – e familiari – la presenza di figli, la cittadinanza dei componenti, il numero di percettori di reddito – nel determinare il rischio di povertà fra i lavoratori dipendenti residenti nei comuni del Nord-Ovest. Nella figura 9.2 è riportata la distribuzione dei lavoratori dipendenti nei quattro gruppi (no *working poor*, lavoratori poveri in famiglie povere, lavoratori poveri in famiglie non povere, lavoratori non poveri in famiglie povere) in funzione delle determinanti individuate col modello di regressione.

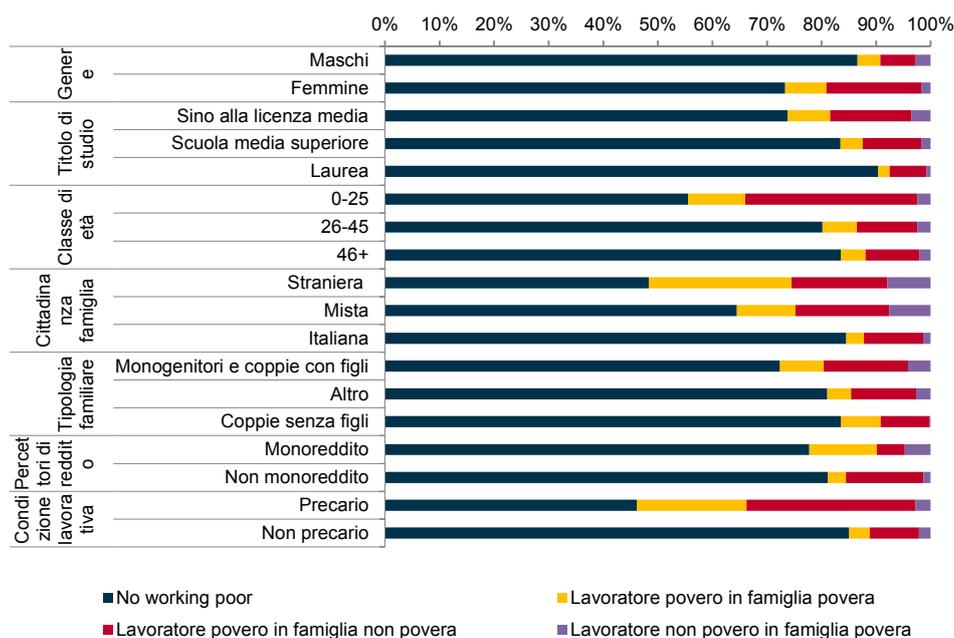
Per i lavoratori poveri che vivono in famiglie non povere, il rischio povertà, derivante dai bassi livelli di reddito individuale, è mitigato dalla presenza nel nucleo familiare di più percettori di reddito. Solo il 12 per cento degli individui del gruppo, infatti, vive in famiglie monoreddito, mentre la tipologia familiare largamente predominante è rappresentata dalla coppia con figli e con almeno due percettori di reddito (47,9 per cento). Il gruppo, composto per la gran parte da cittadini italiani (80 per cento), è caratterizzato da una rilevante incidenza di lavoratori con un titolo di scuola media superiore (43 per cento) e di precari (32,6 per cento), soprattutto donne (22 per cento).

Pur conseguendo redditi da lavoro dipendente superiori alla soglia di povertà, il 2,2 per cento dei lavoratori si trova a vivere in famiglie povere e ciò accade perché, nella gran parte dei casi, sono gli unici percettori di reddito nella famiglia. Infatti, il 57,7 per cento dei lavoratori che rientrano in questo gruppo vive in famiglie monoreddito, di cui il 32 per cento in coppie monoreddito con figli e il 9 per cento in famiglie con un solo genitore. I lavoratori con contratto a tempo indeterminato rappresentano l'84,6 per cento degli individui del gruppo, che si caratterizza anche per una elevata incidenza di lavoratori over 25 anni (94 per cento) con basso livello di istruzione (un lavoratore su due ha più di 25 anni e ha conseguito al massimo il diploma di scuola media inferiore).

Al netto del sottogruppo dei lavoratori che vivono da soli, fra i lavoratori poveri sia a livello individuale che familiare, la quota più rilevante (58,8 per cento) vive in

famiglie con almeno due percettori di reddito. Il gruppo è caratterizzato da una significativa incidenza di precari (43,1 per cento), per lo più stranieri (25,4 per cento), di lavoratori in età compresa fra 25 e 45 anni (56,2 per cento), con bassi livelli di scolarizzazione (il 44,5 per cento ha completato, al massimo, la scuola dell'obbligo).

Figura 9.2 - Lavoratori a rischio di povertà e non per le principali determinanti. Nord-Ovest, 2014



9.2.3 La diffusione territoriale del fenomeno

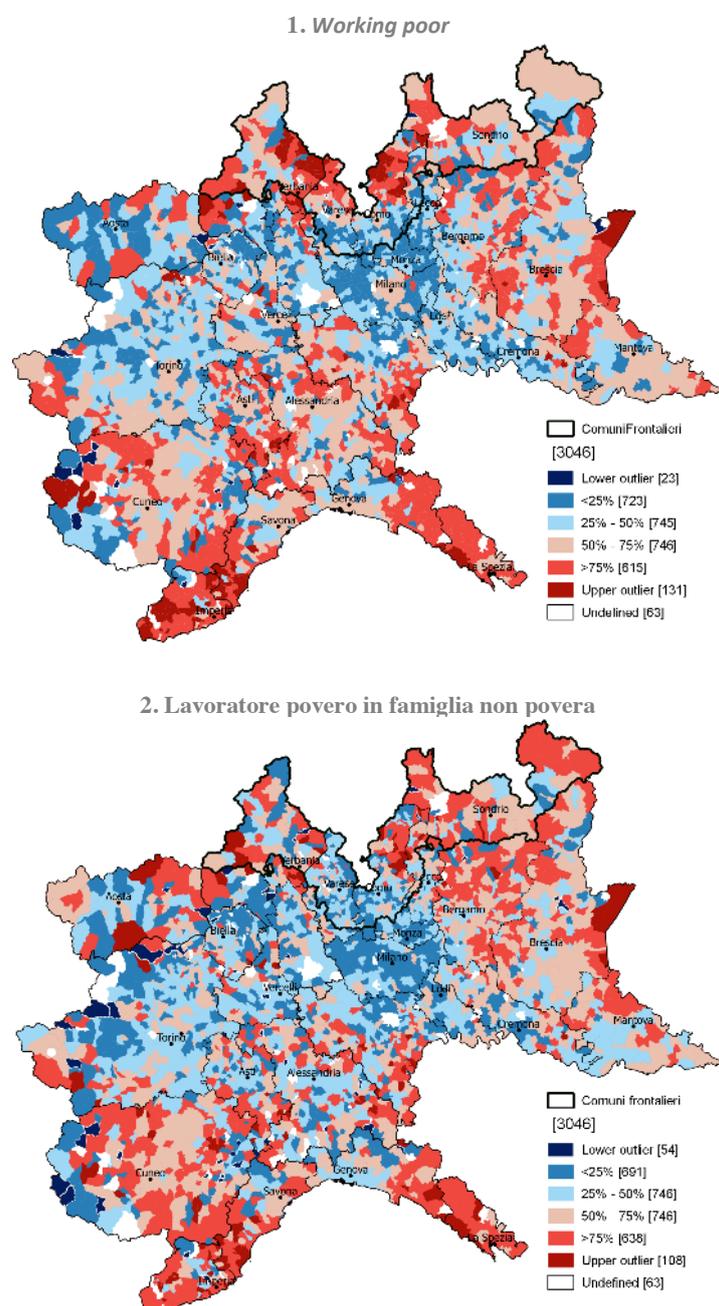
L'analisi dei dati comunali consente di individuare le aree di particolare incidenza del fenomeno. Gran parte dei comuni liguri, alcune aree del basso Piemonte, i comuni della zona del Garda sono aree in cui si un lavoratore dipendente su cinque si posiziona nell'area del *working poor*. In queste stesse aree tendono a concentrarsi i comuni con le percentuali più elevate di lavoratori poveri in famiglie povere e di lavoratori poveri che vivono in famiglie non povere (Cartogrammi 4 e 2 di Figura 9.3); le aree di maggiore incidenza di lavoratori appartenenti al 2° gruppo (non poveri che vivono in famiglie povere) si collocano tra Basso Monferrato e Lomellina, nell'area lombarda che dall'Oltrepò mantovano si spinge verso Nord-ovest fino alla Val Camonica (Cartogramma 3 di Figura 9.3).

Le indicazioni relative alla localizzazione del fenomeno sono confermate anche dalle *cluster map* di figura 9.4, che evidenziano la presenza, in quei territori, di gruppi di comuni contigui caratterizzati da elevate incidenze di *working poor* (cluster *High-High*).

Oltre alle aree sopra richiamate, i cartogrammi evidenziano la presenza del fenomeno anche nell'area di confine con la Svizzera. In questo caso, però, c'è da considerare il possibile effetto distorsivo del fenomeno del lavoro transfrontaliero. Infatti, poiché in virtù dell'accordo Svizzera-Italia i redditi dei lavoratori frontalieri sono dichiarati nel paese in cui sono prodotti, i redditi individuali e familiari pre-

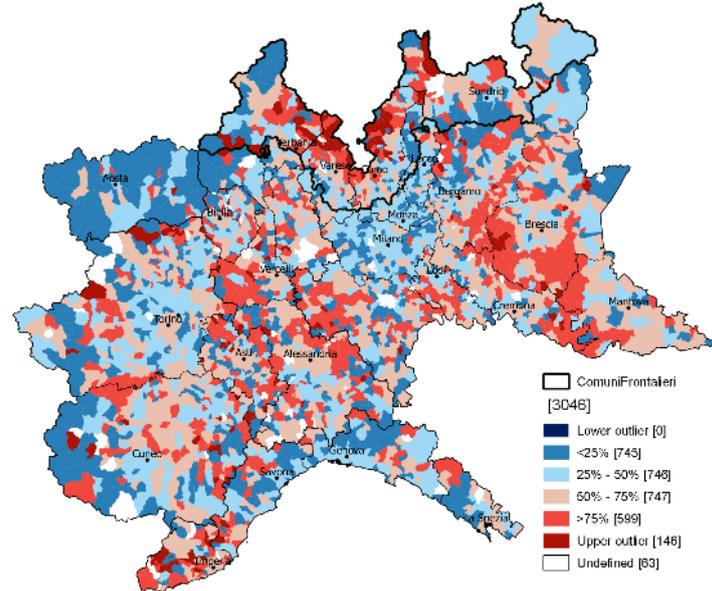
senti nelle basi dati di ArchIMEDe non includono la quota imputabile ad attività lavorative in Svizzera⁵.

Figura 9.3 - Lavoratori a rischio povertà, incidenza per comune. Nord-Ovest, 2014



⁵ Il lavoro oltre confine in Francia (e in Svizzera per i comuni lontani oltre 20 km dal confine) prevede una quota esente nonostante la dichiarazione dei redditi sia prevista nel Paese di residenza del lavoratore (Cfr. Appendice 2).

3. Lavoratore non povero in famiglia povera



4. Lavoratore povero in famiglia povera

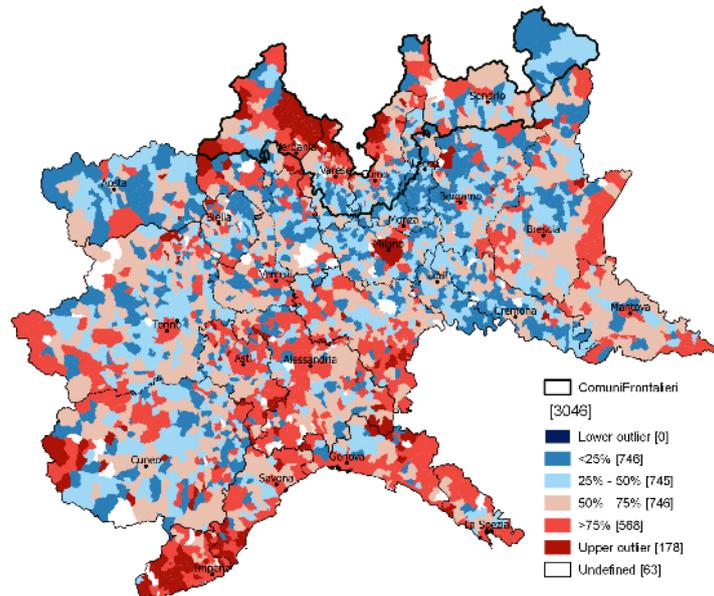
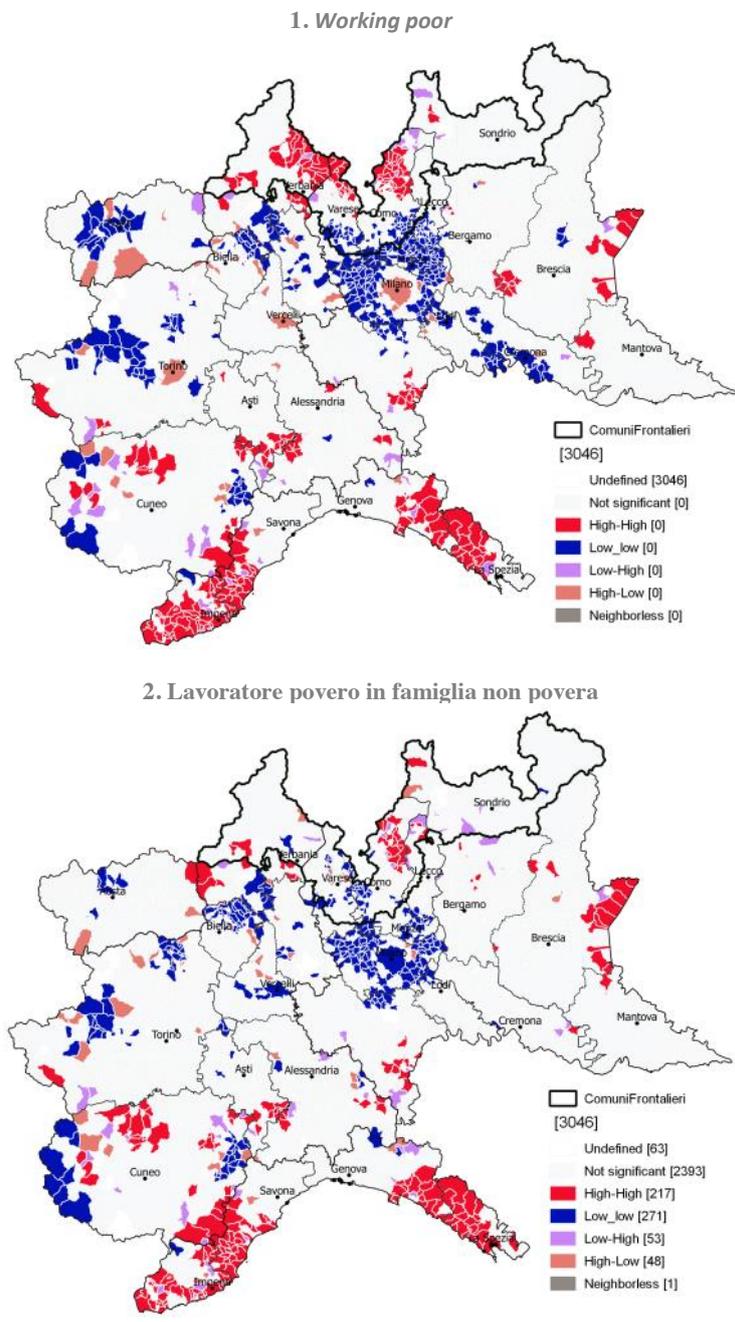
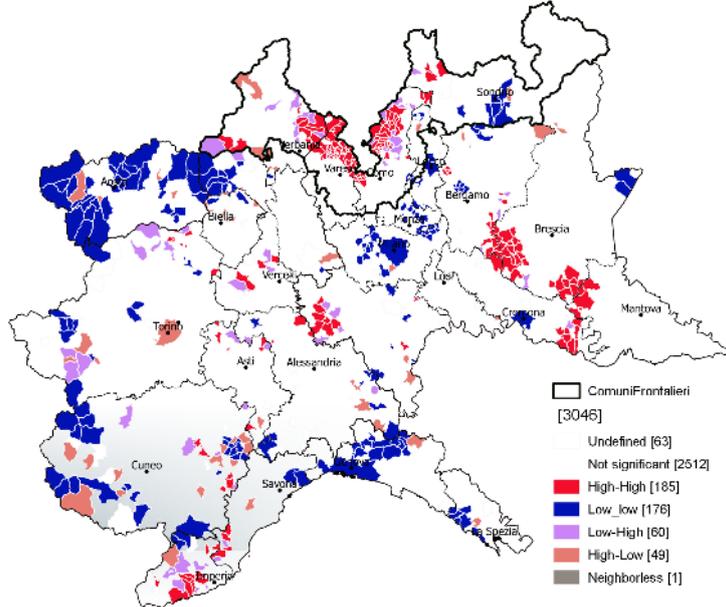


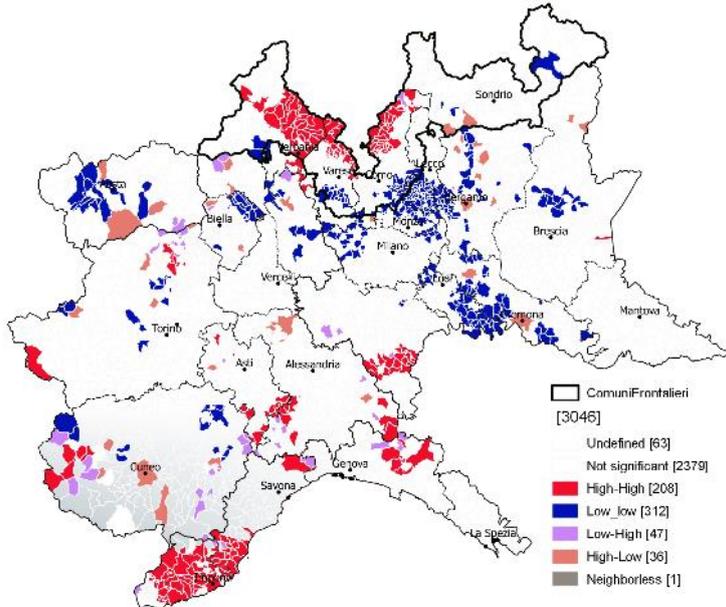
Figura 9.4 - Incidenza lavoratori a rischio povertà. Nord-Ovest, 2014 (Cluster map)



3. Lavoratore non povero in famiglia povera



4. Lavoratore povero in famiglia povera



9.3 Conclusioni

Il lavoro rappresenta una prima verifica della possibilità di utilizzare le basi dati del Progetto ArchIMEDe per una analisi della diffusione territoriale del fenomeno del *working poor*.

I risultati qui sintetizzati evidenziano le potenzialità informative dei *data set* di ArchIMEDe. Difatti il quadro conoscitivo che emerge dall'analisi delle determinanti della povertà condotta sui dati di fonte amministrativa integrati dall'Istat in ArchIMEDe risulta coerente con quello frutto di elaborazioni su dati di indagine e il ricorso a dati di fonte amministrativa, opportunamente integrati, ha consentito di approfondire il livello di dettaglio territoriale dell'analisi, producendo cartografie comunali di incidenza del fenomeno indagato.

CAPITOLO X
PROPOSTA DI UN INDICE
DI VULNERABILITÀ SOCIO-ECONOMICA¹
di Raffaella Succi, Francesca Vannoni e Alberto Vitalini

10.1 Introduzione

Le condizioni sociali ed economiche che caratterizzano gli individui e le famiglie, come i contesti in cui essi vivono, sono un fenomeno di interesse in diversi ambiti, quali quello economico, sociale e sanitario. Già a partire dagli anni '80 in Inghilterra vengono sviluppati indici sintetici di deprivazione sulla base dei dati disponibili dalle rilevazioni censuarie, con l'intento di considerare altre dimensioni oltre a quella reddituale nello studio delle condizioni di povertà relativa (Townsend, 1987). In Italia a partire dagli anni '90, sulla base dei dati del Censimento della popolazione e delle abitazioni, è stato definito un indice di deprivazione nazionale, elaborato prima a livello comunale (Cadum et al., 1999) e poi a livello di sezioni di censimento (Caranci et al., 2009; Caranci et al., 2010). Recentemente l'Istat nell'ambito del progetto 8milaCensus ha fornito uno strumento utile per studiare l'evoluzione delle condizioni sociali e materiali del territorio considerando alcuni degli indicatori elaborati per le varie occasioni censuarie a partire dal censimento del 1956 e sintetizzandole in un indice di vulnerabilità sociale e materiale².

La recente disponibilità di registri statistici ottenuti grazie all'integrazione di archivi amministrativi consente di arricchire l'offerta informativa ad un elevato livello di granularità territoriale e con una periodicità di aggiornamento che supera il limite della cadenza decennale con cui è stato tradizionalmente condotto il censimento della popolazione. Utilizzando queste fonti è possibile elaborare un indice sintetico a partire dai microdati elementari con una cadenza annuale.

L'obiettivo di questo lavoro è quello di presentare i risultati di una prima sperimentazione volta a testare il potenziale informativo degli archivi tematici "Condizioni socio-economiche delle famiglie" e "Precarietà lavorativa" del progetto ArchIMeDe (Garofalo, 2014) per l'elaborazione di un indice sintetico di vulnerabilità socio-economica che considera, come unità statistica di analisi, unità geografiche. In questa sperimentazione, condotta sulle regioni Liguria e Piemonte, vengono considerati i comuni, ma con l'intento di arrivare al dettaglio di sezione o unità urbanistica, per le aree metropolitane, non appena sarà disponibile tale dettaglio informativo.

La perdurante carenza di risorse a disposizione delle politiche pubbliche rende ancora più necessario soddisfare le esigenze informative dei decisori che a vari gradi di responsabilità – da quello nazionale fino a quello locale – hanno il compito di mettere in campo delle azioni, tenendo conto del contesto in cui operano. A tale

¹ Il capitolo è frutto della collaborazione degli autori Raffaella Succi, Francesca Vannoni e Alberto Vitalini. In particolare, il paragrafo 10.2 è da attribuire a Raffaella Succi (succi@istat.it), i paragrafi 10.1, 10.4 e 10.5 sono da attribuire a Francesca Vannoni (vannoni@istat.it), il paragrafo 10.3 è da attribuire ad Alberto Vitalini (vitalini@istat.it). L'elaborazione dei dati mediante SAS ed il software Comics è stata effettuata da Raffaella Succi, mentre le cartografie sono state prodotte da Alberto Vitalini mediante GeoDA (Cfr. Capitolo 4).

² Indice di vulnerabilità sociale e materiale (Istat, 8milaCensus). Online: http://ottomilacensus.istat.it/fileadmin/download/Indice_di_vulnerabilità_sociale_e_materiale.pdf.

scopo servono dati aggiornati che consentano di scendere ben oltre il dettaglio regionale o provinciale, arrivando almeno fino al livello comunale e, per i comuni più grandi, anche sub-comunale.

Gli indici di deprivazione hanno l'obiettivo di rispondere a diverse esigenze di politica pubblica. In ambito sanitario e ambientale servono, da un lato, per orientare gli interventi di contrasto alle disuguaglianze di salute tenendo conto dei determinanti sociali della salute e, dall'altro lato, per sottrarre il potenziale confondimento socio-economico dalle misure d'associazione della salute con i fattori ambientali (Pasetto, 2011). Nel campo delle politiche sociali, questi indici sono indispensabili per la programmazione e valutazione degli interventi di contrasto e prevenzione delle condizioni che portano all'esclusione sociale o che sono implicate nei meccanismi generatori delle disuguaglianze sociali attinenti le diverse sfere della vita delle persone e delle loro famiglie (Rosano et al., 2016; Costa et al., 2011).

10.2 Lo schema teorico di riferimento e il contenuto informativo dell'archivio

La scelta delle variabili da inserire nell'indice sintetico prende come riferimento la definizione di deprivazione di Townsend, sulla quale si basano anche gli indici di deprivazione già sviluppati a livello nazionale per l'analisi delle disuguaglianze sociali di salute (Caranci *et al.*, 2009). Secondo questo schema le dimensioni rilevanti per il fenomeno della deprivazione sono le seguenti: (i) istruzione e formazione; (ii) mercato del lavoro; (iii) risorse materiali e (iv) reti familiari. Il modello impiegato è di tipo formativo, per cui gli indicatori utilizzati per misurare ciascuna dimensione sono visti come causa del fenomeno da misurare (Mazziotta e Pareto, 2015). Per ciascuna dimensione si è considerato l'insieme delle variabili presenti nell'archivio ArchIMEDe candidate ad entrare nell'indice aggregato tramite l'elaborazione di idonei indicatori. Sulla base dell'analisi della rilevanza e accuratezza di questi indicatori, è stato scelto quello risultato più idoneo a rappresentare la dimensione di appartenenza, assumendo come punto di riferimento l'indice di deprivazione nazionale elaborato in ambito epidemiologico (Caranci *et al.*, 2010). Tale attività ha portato ad escludere alcuni indicatori che sul piano teorico sarebbero dovuti essere inclusi, come ad esempio i giovani che non studiano e non lavorano (*Neet*). Gli indicatori utilizzati sono descritti nel prospetto 10.1, dove sono riportati per ciascuna dimensione gli indicatori semplici che compongono l'indice di deprivazione di fonte censuaria e gli indicatori semplici scelti per l'indice di vulnerabilità. Nel prospetto 10.2 sono riportate le definizioni degli indicatori di fonte amministrativa, come concordate nell'ambito di altri progetti Istat che si basano sulle statistiche sperimentali. Si osservi che a differenza degli indici sintetici costruiti sulla base dei dati censuari, il cui contenuto informativo consente di misurare la dimensione "Risorse materiali" attraverso due variabili sulla condizione abitativa, la fonte ArchIMEDe consente di misurare le dimensioni "Condizione lavorativa" e "Risorse materiali" mediante due indicatori, quali la "Bassa intensità lavorativa" e la "Bassa disponibilità di redditi lordi dichiarati", importanti per individuare condizioni a rischio di povertà relativa e di esclusione sociale. Si ritiene dunque che, utilizzando un modello di sintesi formativo, l'indice di vulnerabilità socio-economica basato sui dati di ArchIMEDe possieda una capacità descrittiva differente da quella mostrata dagli indici basati esclusivamente sui dati di fonte censuaria. Ciò è dovuto al fatto che la fonte ArchIMEDe consente di misurare direttamente la capacità reddituale delle famiglie e degli individui mentre ciò non avviene utilizzando i dati del censimento della popolazione. Come si vedrà nel paragrafo successivo, la misura dello svantaggio socio-economico risulta più precisa in alcuni

contesti nei quali le variabili sulla dimensione e sul titolo di godimento dell'abitazione non risultano essere delle buone misure della condizione materiale, come nei comuni rurali/montanti o nei comuni caratterizzati da fenomeni turistici.

Prospetto 10.1 - Indicatori scelti per l'indice di vulnerabilità e confronto con l'indice di deprivazione

| DIMENSIONI | INDICE VULNERABILITA' <i>fonte ArchIMEDe</i> | INDICE DEPRIVAZIONE <i>fonte Censimento</i> |
|-----------------------|--|---|
| Cultura/formazione | Bassa Istruzione | Bassa Istruzione |
| Condizione lavorativa | Bassa intensità lavorativa | Tasso di disoccupazione |
| Risorse materiali | Bassa disponibilità dei redditi lordi dichiarati | Affollamento abitativo Abitazioni in affitto |
| Rete familiare | Famiglie monogenitoriali (con intestatario con meno di 60 anni) con figli dipendenti | Famiglie monogenitoriali con figli dipendenti |

Prospetto 10.2 - Definizioni degli indicatori che compongono l'indice di vulnerabilità socio-economica

| Dimensione | Etichetta | Definizione |
|-----------------------|--|--|
| Cultura/formazione | Bassa Istruzione | Percentuale di persone iscritte in anagrafe di 15 anni e più che: non hanno ancora conseguito la licenza media (fino ai 19 anni); hanno conseguito al più la licenza media (fino ai 59 anni); hanno conseguito al più la licenza elementare (dai 60 anni in poi). |
| Condizione lavorativa | Bassa intensità lavorativa | Percentuale di persone che risiedono in famiglie anagrafiche la cui intensità di lavoro è inferiore a 0,20, ossia incidenza di persone che vivono in famiglie dove le persone in età lavorativa (tra i 18 e i 59 anni, con l'esclusione degli studenti 18-24 anni) hanno lavorato per meno del 20 per cento del loro potenziale (con esclusione delle famiglie composte solo da minori, da studenti di età inferiore a 25 anni e da persone di 60 anni o più). |
| Risorse materiali | Bassa disponibilità dei redditi lordi dichiarati | Percentuale di persone residenti in famiglie anagrafiche con reddito lordo equivalente pro-capite familiare minore del 60 per cento del reddito equivalente mediano nazionale sul totale delle persone residenti in famiglie anagrafiche. |
| Rete familiare | Famiglie monogenitoriali (con intestatario con meno di 60 anni) con figli dipendenti | Percentuale di famiglie anagrafiche monogenitoriali (con intestatario con meno di 60 anni) con figli dipendenti secondo le fonti amministrative del Fisco (non occupati e senza reddito) sul totale delle famiglie anagrafiche. |

Il metodo di sintesi scelto per l'indicatore è quello parzialmente compensativo sviluppato da Mazziotta e Pareto (Mazziotta e Pareto, 2015), metodo coerente con

il modello formativo a variabili non sostituibili³. In particolare è stato utilizzato l'indice Mazziotta-Pareto Index (MPI), poiché l'attenzione è stata concentrata sulla capacità informativa e sulla qualità della fonte utilizzata in un dato anno. Sarà oggetto di un prossimo studio la verifica della qualità dell'archivio ArchIMEDe per le analisi secondo una prospettiva temporale, mediante l'utilizzo dell'*Adjusted Mazziotta Pareto Index* (AMPI), tenendo anche conto che l'AMPI utilizza un metodo di standardizzazione *min-max*, per il quale dunque la presenza di *outlier* incide in modo significativo sul risultato finale.

Anche se disponibili dati più aggiornati, si è scelto di svolgere questa sperimentazione considerando ArchIMEDe per l'anno 2012, in modo da rendere più significativo il confronto con gli indici basati sulla fonte censuaria, il cui riferimento temporale è l'8 ottobre 2011.

10.3 Risultati

Le analisi di sensitività, ottenute mediante il *software Comic* utilizzato per l'elaborazione⁴, confermano la coerenza dell'indice con il quadro teorico generale nella scelta delle variabili, mentre l'analisi di robustezza conferma un'elevata variabilità dell'indice a causa della presenza di comuni con dimensioni molto piccole. Questo limite motiva la scelta di replicare in futuro le analisi per ambiti territoriali differenti, più omogenei in termini di numerosità della popolazione, come ad esempio quelli utilizzati per la sperimentazione della misura di contrasto alla povertà SIA (Sostegno per l'Inclusione Attiva) e rilevanti anche per le politiche socio-sanitarie.

La buona capacità discriminante dell'indice di vulnerabilità socio-economica risulta evidente dalle cartografie riportate di seguito, in cui si osservano aree maggiormente vulnerabili nella Liguria del ponente ed in alcune aree interne dell'area del levante, mentre i comuni della costa risultano quasi tutti non socio-economicamente vulnerabili (Figure 10.1, 10.3). Queste evidenze sono coerenti con i risultati ottenuti utilizzando l'indice di deprivazione nazionale elaborato per la Liguria sulla base dei dati censuari.

L'analisi dell'autocorrelazione spaziale per la regione Liguria (Figura 10.5), evidenzia due aree ben definite: da un canto, quella nell'estremo ponente e nell'entroterra genovese della Valtrebbia, costituita dai comuni che presentano un'elevata vulnerabilità socio-economica e un'alta correlazione con i comuni limitrofi (in rosso); dall'altro canto, nella provincia di Savona e nell'estremo levante, l'area composta dai comuni che presentano basso livello di vulnerabilità e bassa correlazione con i comuni confinanti (in blu). Nell'analisi di questi risultati va introdotto un *caveat*. L'ampiezza e l'intensità della vulnerabilità misurate nella provincia di Imperia potrebbe almeno in parte dipendere dal fenomeno del pendolarismo oltre confine per fini lavorativi⁵. Infatti anche se, per i redditi prodotti in Francia la dichiarazione avviene nel Paese di residenza, e di conseguenza lavoro e redditi sono presenti in ArchIMEDe, esiste una franchigia sull'imponibile. Sull'impatto della franchigia sull'indice di vulnerabilità saranno necessari utili ulteriori approfondimenti.

³ In un modello a variabili non sostituibili l'ipotesi di base è che la variazione di una variabile non può essere compensata dalla variazione in senso contrario di un'altra.

⁴ Massoli P., Mazziotta M., Pareto A., Rinaldelli C. (2015).

⁵ Gli occupati residenti in famiglia che si recano al luogo abituale di lavoro all'estero sono 3.901 per la provincia di Imperia: circa il 6 per cento degli occupati totali residenti nella provincia (Dati Censimento della popolazione 2011).

In Piemonte figurano socio-economicamente vulnerabili le aree lungo l'arco alpino e l'Appennino Ligure, più un'ampia zona a cavallo delle province di Vercelli ed Alessandria, sebbene a macchia di leopardo per la compresenza in questa area di comuni con valori elevati dell'indice accanto a comuni che invece sono in linea con la media regionale (Figure 10.2, 10.4). Come si può osservare nei cartogrammi, sono stati esclusi dall'analisi i comuni della provincia del Verbano-Cusio-Ossola, dove è presente il pendolarismo oltre confine per fini lavorativi⁶ e quindi sarebbero potuti risultare erroneamente i problemi della bassa intensità lavorativa e il rischio di povertà (Cfr. Appendice 2). Nella base dati ArchIMEDe, infatti, non sono disponibili le informazioni sul reddito e sull'occupazione delle persone occupate in Svizzera e, di conseguenza, queste ultime risultano – apparentemente – senza lavoro e quindi a basso reddito.

In Piemonte, oltre a confermare le aree di vulnerabilità già evidenziate dal box plot, l'analisi dell'autocorrelazione (Figura 10.6) permette di individuare un'ampia area con un basso livello di vulnerabilità che si sviluppa idealmente lungo l'asse che unisce la città di Alba al capoluogo di Cuneo.

10.4 Criticità

Questo studio al momento presenta tre limiti, in particolare: i) le unità per le quali è stato elaborato l'indice sono unità statistiche molto disomogenee per numerosità della popolazione, non essendo ancora disponibili al momento dell'elaborazione i dati per aree sub-comunali; ii) i comuni con pochi abitanti presentano un'elevata variabilità degli indicatori e ciò si riflette sulla bassa robustezza dell'indice; iii) l'indice di vulnerabilità non è standardizzato per età, quindi i risultati sono condizionati dalla struttura per età della popolazione. A tali limiti si aggiungono quelli che derivano dalla mancanza nell'archivio ArchIMEDe di alcune variabili di tipo soggettivo fondamentali, come la parziale disponibilità della condizione occupazionale (non risulta possibile individuare gli individui disoccupati o in cerca di occupazione) e le variabili necessarie per ricostruire le condizioni di esclusione sociale, così come rilevate con l'indagine EU-Silc. Inoltre, disponendo esclusivamente della variabile sul reddito lordo, c'è il rischio di distorcere la stima delle famiglie a basso reddito. A tale riguardo si potrebbero effettuare degli approfondimenti sul confronto delle stime del rischio di povertà ricavate da ArchIMEDe con quelle ottenute coi dati dell'Indagine Eu-Silc per individuare la direzione di questa distorsione. Infine, per i comuni con una bassa numerosità di popolazione le stime di rischio di vulnerabilità socio-economica sono meno precise.

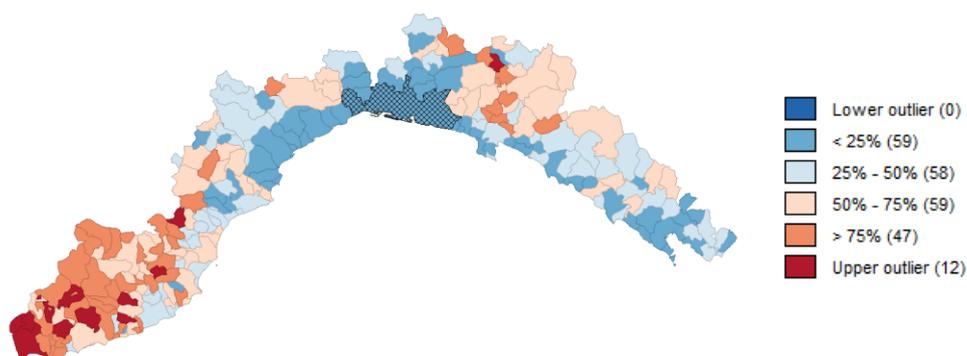
10.5 Sviluppi futuri

Il presente lavoro si presta ad ulteriori approfondimenti, fra cui i) l'analisi della correlazione con la mortalità per l'utilizzo dell'indice di deprivazione a fini epidemiologici; ii) l'impiego, come livello geografico di disaggregazione, degli ambiti territoriali utilizzati per la sperimentazione della misura di contrasto alla povertà SIA e coincidenti per lo più con i distretti sanitari, ma non completamente sovrapponibili a questi ultimi; iii) la costruzione dell'indice a livello individuale per superare i limiti del *bias* dovuto al cambiamento della scala di aggregazione delle unità

⁶ Gli occupati residenti in famiglia che si recano al luogo abituale di lavoro all'estero sono 5.404 per la provincia di Verbano-Cusio-Ossola: quasi il 10 per cento degli occupati totali residenti nella provincia (Dati Censimento della popolazione 2011).

territoriali e iv) il controllo della correlazione di quest'ultimo con la mortalità. L'ipotesi è che l'associazione dell'indice di vulnerabilità socio-economica calcolato a livello individuale con la mortalità risulti più accentuata rispetto a quanto accada con l'indice definito a livello aggregato (comunale, sub comunale o di ambito territoriale), così come già rilevato negli studi di validazione dell'indice di Caranci (Costa et al., 2011; Caranci et al., 2009).

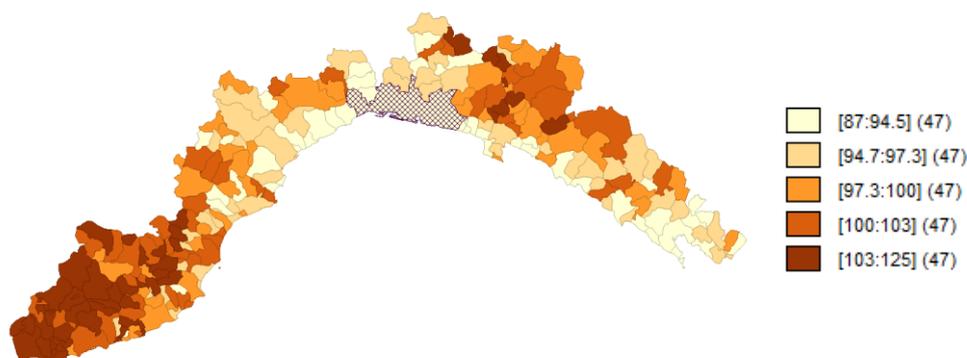
Figura 10.1 - Indice di vulnerabilità socio-economica. Liguria, 2012 (Box plot map)



Fonte: elaborazione su dati ArchIMEDe.

Note: I comuni con almeno 100mila residenti (i.e. Genova) sono evidenziati con la retinatura.

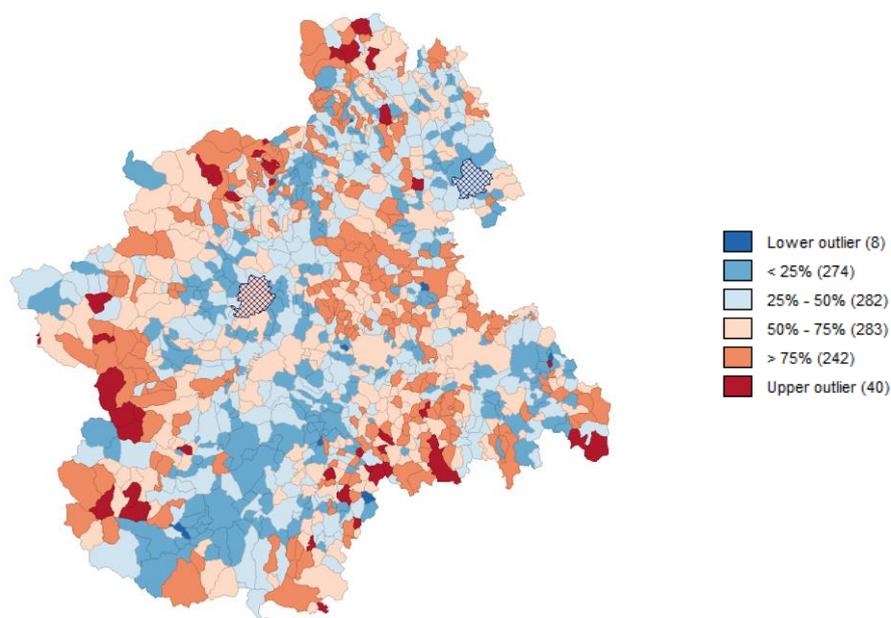
Figura 10.2 - Indice di vulnerabilità socio-economica. Liguria, 2012 (quintili di area)



Fonte: elaborazione su dati ArchIMEDe.

Note: I comuni con almeno 100mila residenti (i.e. Genova) sono evidenziati con la retinatura.

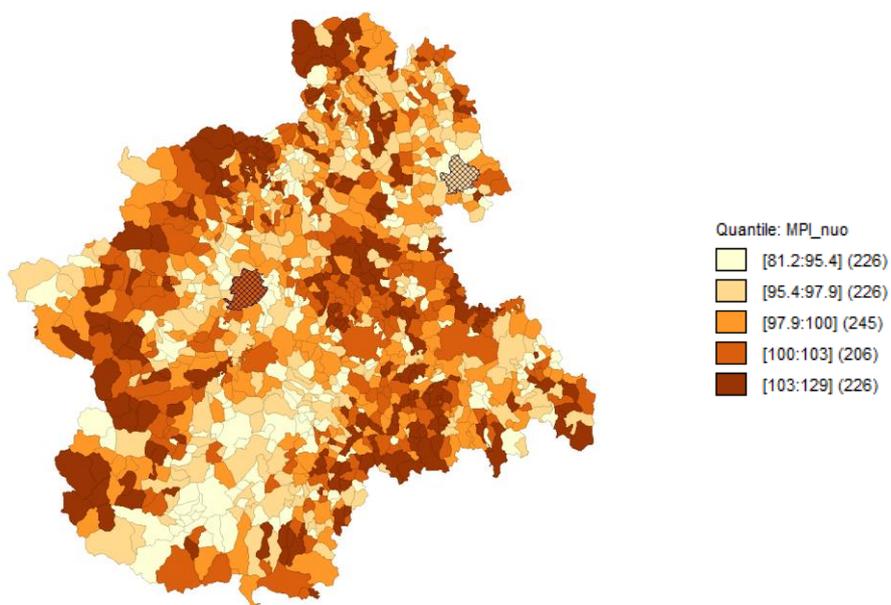
Figura 10.3 - Indice di vulnerabilità socio-economica. Piemonte (senza la provincia di Verbano-Cusio-Ossola), 2012 (Box-plot map)



Fonte: elaborazione su dati ArchIMEDe.

Note: I comuni con almeno 100mila residenti (i.e. Torino, Novara) sono evidenziati con la retinatura.

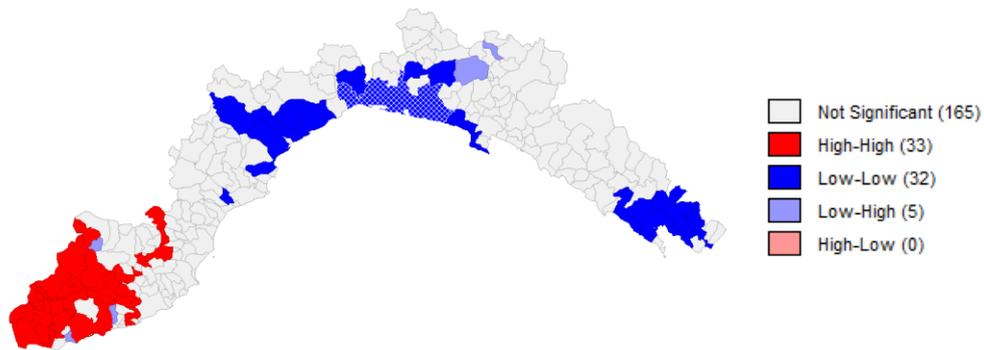
Figura 10.4 - Indice di vulnerabilità socio-economica. Piemonte (senza la provincia di Verbano-Cusio-Ossola), 2012 (quintili di area)



Fonte: elaborazione su dati ArchIMEDe.

Note: I comuni con almeno 100mila residenti (i.e. Torino, Novara) sono evidenziati con la retinatura.

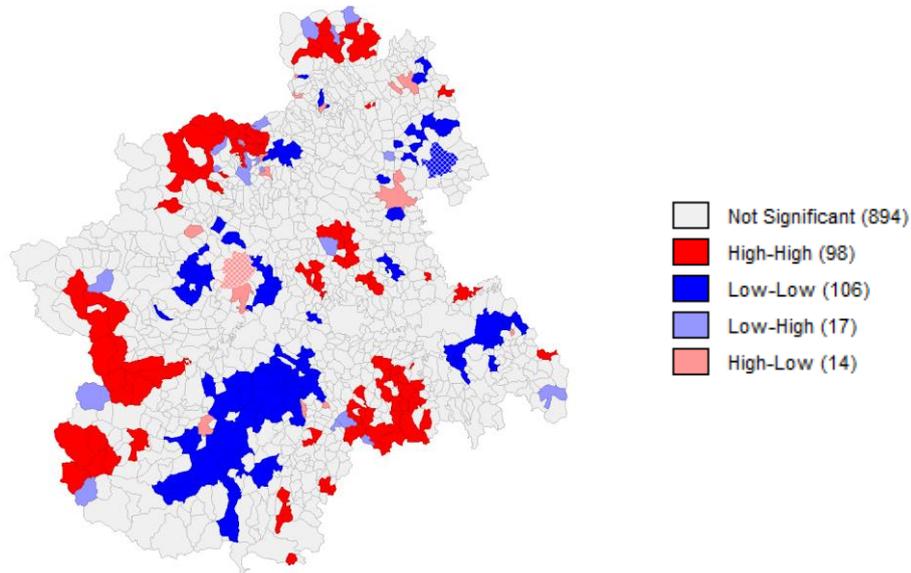
Figura 10.5 - Indice di vulnerabilità socio-economica. Liguria, 2012 (LISA cluster map)



Fonte: elaborazione su dati ArchIMEDe.

Note: I comuni con almeno 100mila residenti (i.e. Genova) sono evidenziati con la retinatura.

Figura 10.6 - Indice di vulnerabilità socio-economica. Piemonte (senza la provincia di Verbano-Cusio-Ossola), 2012 (LISA cluster map)



Fonte: elaborazione su dati ArchIMEDe.

Note: I comuni con almeno 100mila residenti (i.e. Torino, Novara) sono evidenziati con la retinatura.

CAPITOLO XI
REDDITI DELLE FAMIGLIE E
CONDIZIONI DI FRAGILITÀ IN LOMBARDIA
di Sara Casacci

11.1 Premessa

Un'ampia letteratura italiana e internazionale ha affrontato il tema del benessere materiale di individui e famiglie, specialmente dopo che la crisi economica ha prodotto effetti negativi sulle condizioni di vita delle persone. In tale contesto, assume notevole rilevanza la disponibilità di indicatori riconducibili al reddito, al patrimonio e ai consumi delle famiglie, esplicitamente richiamati dall'Istat nel progetto sul Benessere Equo e Sostenibile (Istat, 2016).

Per quanto concerne il reddito, che rappresenta uno dei principali fattori che descrivono il benessere economico, lo studio dei livelli e della disuguaglianza nella sua distribuzione, unitamente a quello del rischio di povertà, costituisce un filone molto sviluppato. In Italia l'indagine sul reddito e le condizioni di vita (Eu-Silc) dell'Istat rappresenta una delle fonti principali per la conoscenza delle condizioni economiche di individui e famiglie: da essa sono derivati, tra gli altri, gli indicatori ufficiali sui redditi e sul rischio di povertà, anche in funzione delle principali caratteristiche delle famiglie (numero di componenti, numero di percettori di reddito, tipologia familiare, presenza di minori o di anziani). Dal momento che questi dati, derivanti da indagini campionarie, consentono stime per domini territoriali fino al livello regionale (Nuts2), molte applicazioni sul tema della povertà ad un livello territoriale più basso sono state effettuate utilizzando stime per piccole aree (Pratesi, 2016; Quintano et al., 2007); esse sono spesso alla base della realizzazione delle cosiddette *poverty maps*, utili per progettare politiche e interventi pubblici (Bedi et al., 2007). Alcuni uffici di statistica utilizzano i dati fiscali per ricavare informazioni sui redditi ad un livello di granularità territoriale inferiore di quanto consentano le stime dei dati da indagine (p.es. in Italia cfr. *Éupolis Lombardia 2015*, Primi, 2015, Comune di Roma, 2016). I dati fiscali tuttavia hanno come unità di riferimento i singoli individui (persone fisiche contribuenti) e non le famiglie. Alcune applicazioni basate sull'integrazione di dati anagrafici con dati fiscali sono state effettuate per il comune di Brescia (Comune, 2014) e di Milano (Poggi e Verrecchia, 2004).

Il presente contributo si propone di illustrare la potenzialità degli archivi amministrativi – ed in particolare dei segnali di reddito – a supporto delle politiche di sostegno destinate alle famiglie. Uno degli aspetti caratterizzanti il contributo è la disaggregazione delle informazioni ad un livello territoriale comunale o provinciale, mentre le fonti tradizionali di informazioni sul tema forniscono stime per domini territoriali meno fini. La disponibilità di informazioni a livello comunale, sebbene non corrispondenti alle definizioni e classificazioni ufficiali, può rappresentare uno strumento di supporto alla programmazione di eventuali interventi pubblici a sostegno delle famiglie. Attraverso i dati è possibile mettere in luce l'eventuale presenza di associazione spaziale dei livelli di reddito e dei divari nei redditi. Infat-

ti, l'esistenza di segregazione spaziale¹ del reddito può ostacolare l'accesso a opportunità e servizi per alcuni segmenti della popolazione. Inoltre, gli indicatori inerenti al reddito e ai divari nei redditi vengono messi in relazione con altri aspetti demografici dei comuni sui quali l'analisi è condotta.

11.2 Dati e metodo

I dati utilizzati sono derivati dalla base dati “Condizioni socio-economiche delle famiglie – ArchIMEDe”. I principali indicatori in esame sono il reddito mediano lordo equivalente (la scala di equivalenza utilizzata è la scala Ocse modificata), al fine di valutare i livelli di reddito delle famiglie nei diversi territori presi in considerazione e il divario tra i redditi delle famiglie (ovvero il rapporto tra reddito equivalente totale al lordo delle imposte posseduto dal 20 per cento delle famiglie iscritte in anagrafe con reddito più alto e quello posseduto dal 20 per cento con reddito più basso). Gli strumenti di *data visualization*, quali le *quartile map*, le *box plot map* e le *cluster map*, consentono di rappresentare la distribuzione territoriale di questi fenomeni e di mostrare l'esistenza di eventuali associazioni spaziali. Nella lettura dei risultati è opportuno tenere presente alcune limitazioni (Cfr. Capitolo 3): la definizione di reddito presenta dei disallineamenti rispetto alla definizione ufficiale, risentendo quindi di una diversa quantificazione, specialmente per alcune voci. Inoltre, alcune voci di reddito non sono disponibili nelle fonti amministrative di input (specialmente alcuni redditi esenti, a tassazione separata o soggetti ad imposta sostitutiva); è il caso anche dei redditi percepiti dai lavoratori italiani frontalieri per i quali vige la podestà impositiva esclusiva del paese in cui i redditi sono prodotti. Per questo motivo sono stati esclusi dall'analisi i comuni delle province di Varese, Como e Sondrio a causa di una carenza informativa sui redditi percepiti dai lavoratori frontalieri². Un'ulteriore limitazione è costituita dal fatto che le informazioni reddituali fanno riferimento al reddito al lordo delle imposte e non risentono quindi del loro effetto redistributivo.

11.3 Risultati

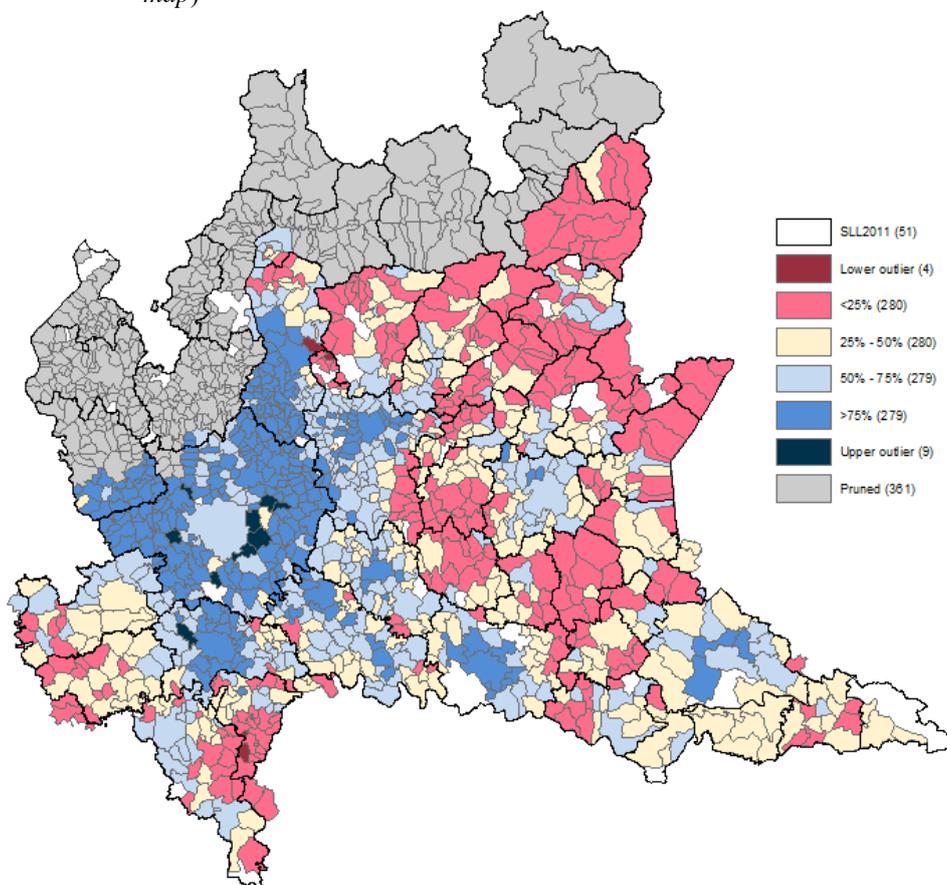
11.3.1 I livelli di reddito delle famiglie lombarde

Dalla *box-plot map* del reddito mediano equivalente familiare (Figura 11.1) emerge una configurazione abbastanza netta dei comuni della Lombardia. I comuni appartenenti all'ultimo quartile della distribuzione territoriale (quelli in cui il reddito familiare risulta più elevato) sono concentrati nelle province di Milano e Monza e Brianza. I comuni che ricadono nel primo quartile della distribuzione territoriale del reddito delle famiglie sono invece concentrati prevalentemente nelle province di Bergamo e Brescia, che quindi si configurano come “area calda”.

¹ Il fenomeno della segregazione si verifica quando due o più gruppi occupano spazi diversi all'interno della stessa città, regione o stato (Gregory et al., 2009). In questo caso, la segregazione spaziale esiste quando alcune aree mostrano una sovra-rappresentazione di popolazione con un reddito elevato e altre ne mostrano una sotto-rappresentazione.

² I lavoratori frontalieri italiani in Svizzera sono circa 65 mila nel 2013. Di essi una buona quota è residente nelle province di Varese e Como (Comune et al., 2017; Ministero dell'Economia e delle Finanze, 2015).

Figura 11.1 - Reddito mediano familiare equivalente lordo. Lombardia, 2013 (Box-plot map)



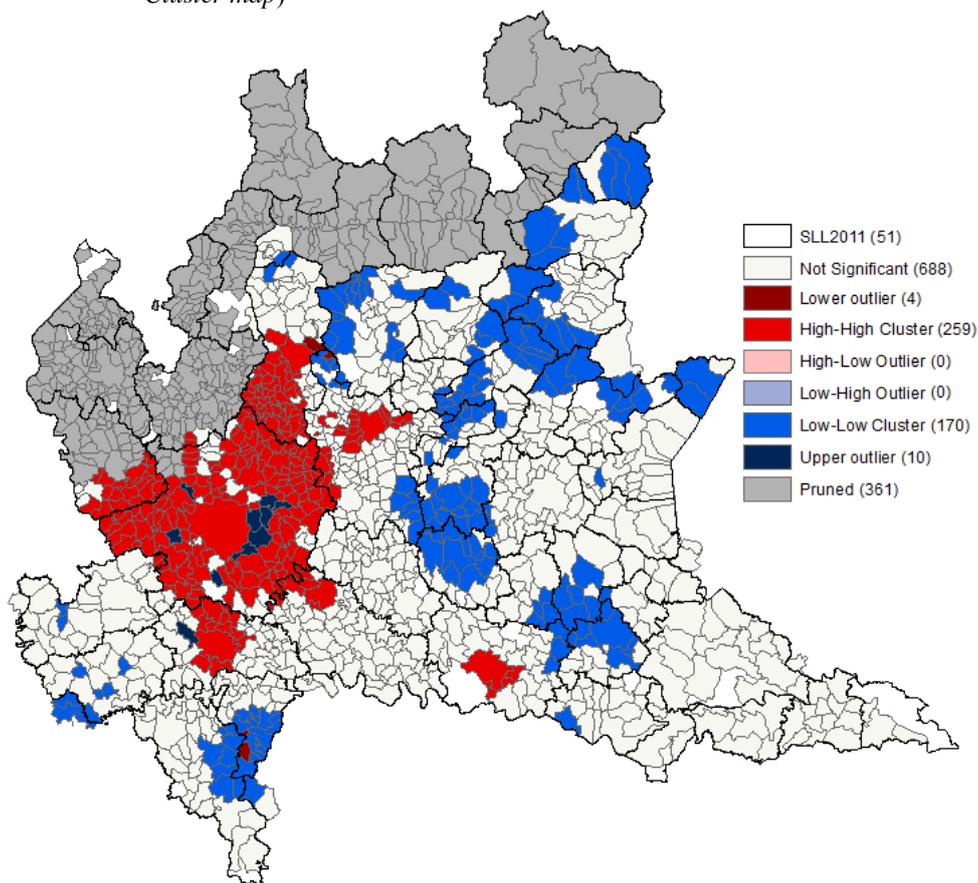
Considerando i Sistemi Locali del Lavoro (SLL), definiti dall'Istat secondo i dati del Censimento della popolazione e delle abitazioni 2011 (Istat, 2015a), è possibile effettuare delle riflessioni tenendo presente la loro specializzazione produttiva³. In primo luogo, si osserva che i comuni del SLL di Milano, classificato come sistema locale urbano ad alta specializzazione, sono quasi tutti appartenenti all'ultimo quartile. È possibile inoltre individuare una fascia di comuni, quella dei SLL di Lodi, Crema e Bergamo, che appartengono prevalentemente al terzo quartile della distribuzione. Questi SLL sono classificati come sistemi locali della manifattura pesante, in particolare della petrolchimica e della farmaceutica. I SLL di Pavia, classificato come sistema locale urbano pluri-specializzato, e di Cremona, con specializzazione produttiva prevalente relativa alla produzione e lavorazione dei metalli, presentano un numero elevato di comuni appartenenti al primo quartile.

La *cluster map* (Figura 11.2) individua dei cluster, anche di dimensione estesa. Il cluster con la maggiore estensione territoriale è di tipo *high-high*: i comuni di questo tipo presentano alti livelli di reddito e comuni vicini con alti livelli di reddito. Il cluster include un'ampia quota dei comuni dell'area metropolitana di Milano e i comuni limitrofi appartenenti alle province di Lecco, Monza e Brianza e Pavia. Questi comuni individuano un'area diffusa in cui le famiglie sono in grado di produrre alti livelli di reddito. I cluster di tipo *low-low* (bassi livelli di reddito e comuni vicini con bassi livelli di reddito) presentano una minore estensione territoriale

³ Fonte: Istat, Classificazioni dei SL utilizzate per il Rapporto annuale 2015. <http://www.istat.it/it/files/2015/05/ClassSL.zip>.

ma risultano più numerosi. Sono prevalentemente presenti nella provincia di Brescia. Anche alcuni comuni dell'Oltrepò pavese confinanti con la provincia di Piacenza si configurano come cluster *low-low*.

Figura 11.2 - Reddito medio familiare equivalente lordo. Lombardia, 2013 (LISA Cluster map)



11.3.2 Reddito e ampiezza demografica dei comuni

Gli indicatori inerenti al reddito possono essere messi in relazione con altri aspetti demografici dei comuni, come l'ampiezza demografica. La letteratura indica il reddito familiare come uno dei fattori correlati positivamente con l'ampiezza demografica delle città (Boulant et al., 2016) e propone diverse ipotesi esplicative per questo fenomeno. In primo luogo, Behrens e Robert-Nicoud (2014) sostengono che gli individui di maggior talento tendono a muoversi verso le città più grandi, nelle quali vengono pagati stipendi più elevati. In secondo luogo, le aree metropolitane beneficerebbero di economie di agglomerazione che consentirebbero maggiore produttività e retribuzioni più elevate (Duranton e Puga, 2004). Infine, nelle città con elevate specializzazioni (p.es. finanza, *information technology*, manifattura avanzata) i salari sarebbero generalmente più elevati (Florida e Mellander, 2016).

Considerando i comuni lombardi con più di 30 mila persone iscritte in anagrafe⁴, è possibile mettere in relazione l'ammontare della popolazione con il valore

⁴ Ad eccezione dei comuni di Varese, Como e Sondrio.

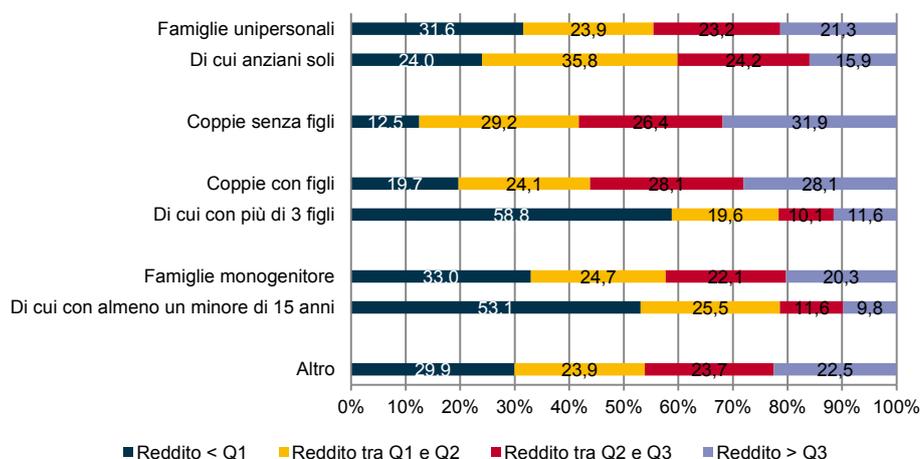
assunto dal reddito mediano lordo delle famiglie da un lato e con il divario nei redditi dall'altro. Mentre tra ampiezza demografica e reddito non sembra esserci alcuna relazione, il divario nei redditi lordi sembra aumentare in funzione dell'ammontare della popolazione iscritta in anagrafe (*r quadro* pari a 0,77). Questo risultato appare in linea con una tendenza analoga, già evidenziata in letteratura (Boulant et al., 2016), per le aree metropolitane dei Paesi Ocse. La relazione, tuttavia, risulta molto rafforzata dalla presenza del comune di Milano, il più popoloso della Lombardia (infatti il valore dell'*r quadro* calcolato senza considerare il comune di Milano è pari a 0,4).

11.3.3 Reddito e caratteristiche delle famiglie

Come evidenziato da numerosi studi e ricerche (Istat, 2011; Istat, 2015b) la composizione delle famiglie – in termini di numerosità, tipologia, età e cittadinanza dei componenti – è un elemento che incide sulla disponibilità di reddito delle famiglie e sul loro rischio di povertà o esclusione sociale. Le politiche territoriali di sostegno alle famiglie sono in genere rivolte a *target* specifici di popolazione, al fine di favorire e sostenere il benessere sociale delle persone e delle famiglie. Esse individuano categorie di beneficiari in funzione di alcune caratteristiche socio-demografiche ed economiche del nucleo familiare quali la tipologia, la presenza di minori, la numerosità, il reddito, ecc.

Le famiglie lombarde⁵ sono state divise in quattro gruppi in base al valore del reddito equivalente percepito nell'anno, utilizzando come soglie i quartili del reddito di tutte le famiglie. Analizzando la distribuzione delle diverse tipologie familiari in questi gruppi (Figura 11.3), si osserva che circa un terzo delle famiglie monogenitore e delle famiglie unipersonali appartengono al primo quartile (rispettivamente 33 per cento e 31,6 per cento). Le coppie senza figli, invece, presentano una quota consistente (31,9 per cento) di famiglie che appartengono all'ultimo quartile, mostrando, quindi, tendenzialmente, un livello di reddito più alto.

Figura 11.3 - Famiglie per tipologia familiare e quartili. Lombardia, 2013

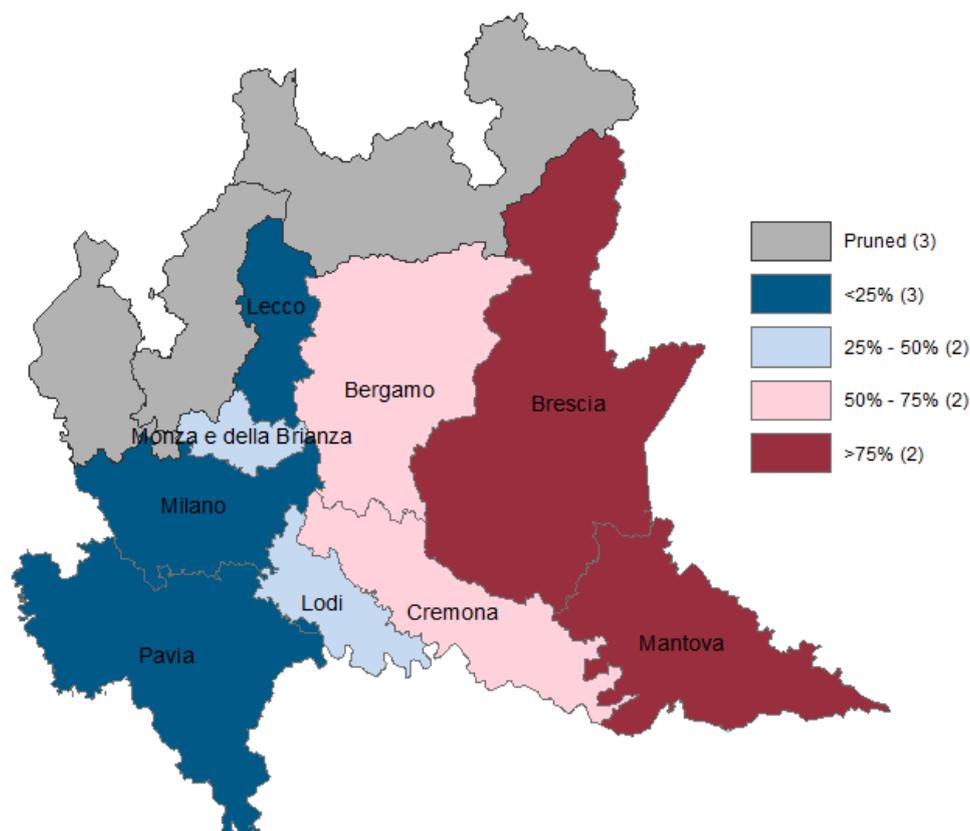


⁵ Con l'esclusione delle famiglie residenti nelle province di Varese, Como e Sondrio.

11.3.4 Gli over 65 soli

Gli anziani con più di 65 anni che vivono da soli, un segmento della popolazione generalmente più a rischio di incorrere in situazioni di riduzione di autonomia e difficoltà di cura della persona (Istat, 2015c), appartengono al primo quartile di reddito in misura minore rispetto al totale delle famiglie unipersonali (24 per cento rispetto al 31,6 per cento del totale delle famiglie composte da una sola persona) e sono maggiormente concentrate nei gruppi intermedi, mostrando dunque di essere, nel complesso, sufficientemente ‘protette’ dal punto di vista economico (Figura 11.3). A livello territoriale, le province di Mantova e Brescia sono quelle in cui si registra per gli anziani soli una maggiore incidenza nel primo quartile di reddito (Figura 11.4). Pavia, Milano e Lecco sono invece le province in cui gli anziani soli sono meno vulnerabili dal punto di vista economico.

Figura 11.4 - Incidenza di anziani soli nel primo quartile del reddito regionale. Lombardia, 2013 (Box-plot map)



11.3.5 Le coppie con figli e le famiglie monogenitoriali.

Per quanto concerne le coppie con figli, dalla Figura 11.3 si evince che le famiglie numerose (con più di tre figli) ricadono per oltre il 58 per cento nel primo quartile. Osservando la distribuzione territoriale (Figura 11.5), si evidenzia una situazione di criticità (in termini di alta incidenza percentuale di coppie con più di tre figli nel primo quartile del reddito regionale) nelle province di Pavia e Mantova.

Milano, Lecco e Monza e Brianza, invece, sono le province in cui le coppie con più di tre figli appaiono meno svantaggiate in termini di reddito.

Infine, relativamente alle famiglie monogenitore, la presenza di almeno un minore di 15 anni rappresenta un fattore di criticità per le famiglie: se la percentuale di famiglie monogenitore presenti nel primo quartile è pari al 33 per cento, l'analoga percentuale per quelle con almeno un figlio under 15 sale al 53,1 per cento. Relativamente alle famiglie monogenitore con almeno un minore di 15 anni, Brescia e Cremona registrano la più alta incidenza nel primo quartile (Figura 11.6), mentre le famiglie di questo tipo residenti nelle province di Milano, Lecco e Monza e Brianza presentano meno frequentemente criticità legate al reddito.

Figura 11.5 - Incidenza di coppie con più di tre figli nel primo quartile del reddito regionale. Lombardia, 2013 (Box-plot map)

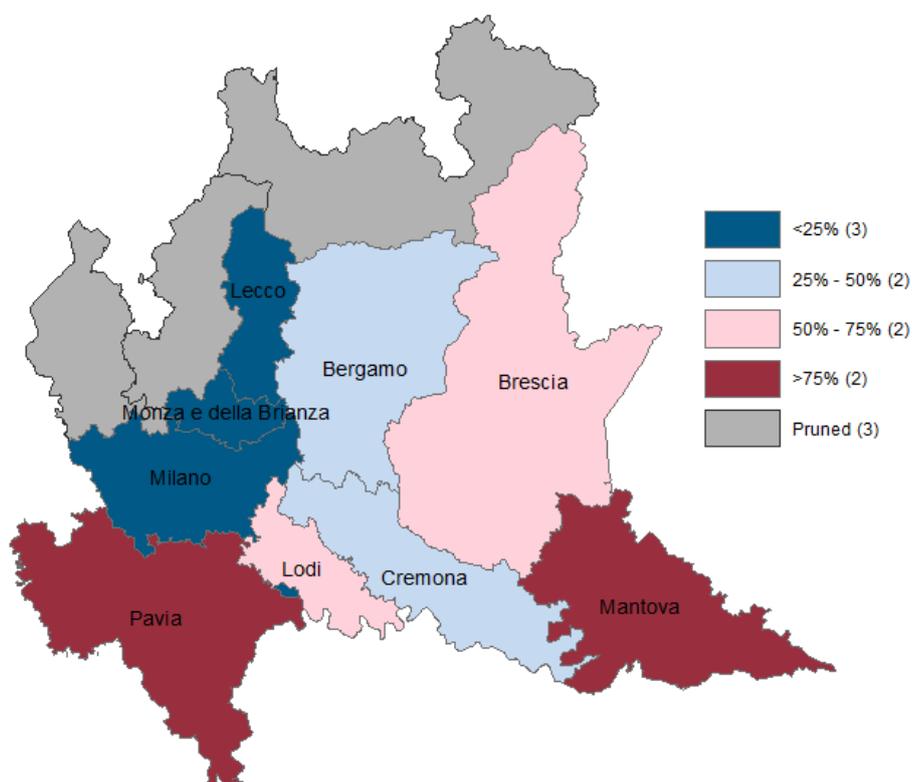
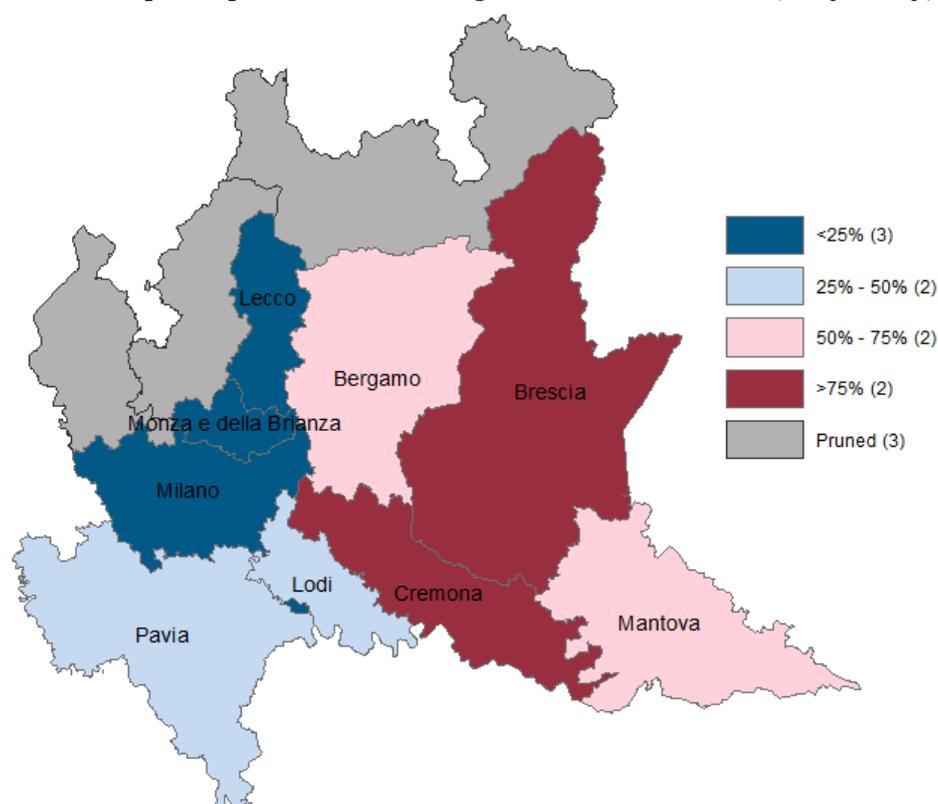


Figura 11.6 - Incidenza di famiglie monogenitore con almeno un minore di 15 anni nel primo quartile del reddito regionale. Lombardia, 2013 (Box-plot map)



11.4 Conclusioni

Il presente lavoro – pur con i limiti segnalati nei capitoli 3 e 9 – ha tentato di evidenziare potenzialità dell’analisi territoriale dei livelli di reddito e dei divari nei redditi delle famiglie lombarde. L’analisi dei dati ha messo in luce la presenza di alcuni territori “critici”. Inoltre, lo studio ha mostrato che alcune tipologie familiari presentano una vulnerabilità economica maggiore rispetto ad altre e si configurano come un target specifico a cui destinare eventuali interventi pubblici. Pertanto, pur mantenendo la possibilità di integrazione con informazioni provenienti da altre fonti, la disponibilità di informazioni ad un livello di granularità territoriale provinciale e comunale può costituire un prezioso strumento di supporto alla programmazione delle politiche locali.

CAPITOLO XII
DISPARITÀ DI GENERE
NEL MERCATO DEL LAVORO IN LOMBARDIA
di Assunta Lisa Carulli, Lorella Sicuro e Domenico Tucci

12.1 Introduzione

In Europa, nel corso del tempo, le donne hanno ottenuto diritti fondamentali per la parità di genere (p.es. possibilità di iscriversi a scuola, lavorare, essere elette, avere posizioni di leadership, divorziare, ecc.), tuttavia, forti disparità sussistono ancora nonostante l'uguaglianza tra uomo e donna sia, oggi, un principio fondamentale, sancito dagli articoli 2 e 3 del Trattato dell'Unione europea e dall'art. 21 della Carta dei diritti fondamentali (European Commission Gender Equality, 2017).

In Italia, lo sviluppo delle politiche di pari opportunità è stato avviato con notevole ritardo rispetto ad altri Paesi europei (p.es. l'accesso all'istruzione e la parità salariale trovano un riconoscimento generalizzato solo dopo la seconda guerra mondiale, grazie alla Costituzione che sancisce formalmente la parità di genere) (World Economic Forum, 2016). Nella seconda metà degli anni settanta, con la Legge 903/77, viene data maggior enfasi agli aspetti lavorativi (p.es. divieto di discriminazione nell'accesso al lavoro, nella formazione professionale, nelle retribuzioni, nell'attribuzione delle qualifiche, delle carriere professionali). Negli anni Ottanta vengono promulgate leggi e creati organismi volti a rafforzare il ruolo sociale delle donne, a promuoverne la partecipazione nel mercato del lavoro e il coinvolgimento attivo nei processi decisionali (Marcucci e Vangelisti, 2013). Tuttavia, nonostante le donne riescano a ricoprire posizioni di maggior rilievo rispetto al passato e, anche in termini quantitativi, abbiano raggiunto livelli occupazionali mai registrati prima, le loro carriere sono spesso interrotte e/o con paghe inferiori, da cui derivano pensioni più basse. I divari di genere nell'occupazione sono particolarmente pronunciati nelle coppie con figli, poiché maggiore è la probabilità che siano le donne a farsi carico delle responsabilità di cura dei figli, riscontrando difficoltà nel conciliare il lavoro con la famiglia.

Il presente contributo, circoscritto alla regione Lombardia, si propone di studiare il divario occupazionale e di reddito tra maschi e femmine e tra donne con e senza figli, a supporto delle politiche pubbliche destinate all'individuo e alla famiglia. A tale scopo, è risultato fondamentale fare riferimento ai dati di fonte amministrativa disaggregati a livello comunale e provinciale, che accanto ai dati di statistica ufficiale costituiscono un valido supporto per focus locali.

12.2 Materiali e metodi

I dati utilizzati riferiti all'anno 2012 derivano dalla base dati Istat "*Condizioni socio-economiche delle famiglie – ArchIMeDe*", dati elementari funzionali alla ricerca sociale ed economica, alla programmazione territoriale e settoriale, nonché alla valutazione delle politiche pubbliche.

Dapprima, sono stati elaborati i principali indicatori quali la dotazione occupazionale e il reddito medio per sesso, calcolati per livello di istruzione e classe di

età¹ (25-34, 35-44, 45-54, 55-64 anni), al fine di valutare i divari di occupazione e di reddito tra maschi e femmine nelle province e nei comuni della regione Lombardia. In particolare, il livello di istruzione include l'istruzione primaria (scuole elementari, medie e diploma di 2-3 anni), l'istruzione secondaria (volta al conseguimento del diploma di 4-5 anni) e l'istruzione universitaria (la laurea e oltre); la condizione lavorativa principale comprende i lavoratori con contratto a tempo indeterminato, a tempo determinato e gli autonomi; il reddito al lordo delle imposte è inteso come il reddito complessivo (al lordo della tassazione). Successivamente, sono stati elaborati diagrammi a barre e utilizzati strumenti di rappresentazione cartografica quali le *box plot map* e le *Lisa cluster map* per esaminare la distribuzione territoriale del fenomeno e mostrare l'esistenza di eventuali similarità spaziali tra ciascuna area e i territori circostanti.

Infine, per studiare le donne con figli rispetto a quelle senza figli riguardo alla possibilità di accedere al mercato del lavoro, sono state svolte analisi statistiche descrittive osservando la quota di donne occupate e non, per titolo di studio, classe di età e tipologia familiare. In riferimento alle stesse variabili si è, infine, classificato il reddito per quartili focalizzando l'attenzione sul primo e sul quarto quartile.

12.3 Caso di studio e statistiche ufficiali

Secondo l'indagine Istat sulle Forze di Lavoro, nell'anno di studio 2012, in Lombardia il 36,4 per cento delle donne e il 43,6 per cento degli uomini hanno un livello d'istruzione primaria; simile invece è il livello d'istruzione secondaria (rispettivamente il 43,0 per cento delle donne e il 42,2 per cento degli uomini), mentre il 17,5 per cento delle donne e il 15,0 per cento degli uomini hanno conseguito un titolo di studio universitario. Sebbene quindi si osservi una quota maggiore di donne con istruzione universitaria (rispetto agli uomini), il tasso di occupazione delle laureate è pari al 78,4 per cento contro l'86,2 per cento di quello maschile (Istat, 2017).

In quanto ai settori di attività economica, i dati Istat dei Censimenti delle imprese e delle istituzioni pubbliche mettono in evidenza che nel 2011 le donne sono impiegate in misura maggiore rispetto agli uomini nel comparto sanità e assistenza sociale e nell'istruzione, con valori rispettivamente pari al 68,1 per cento e al 57,2 per cento. Nel settore privato, la componente femminile prevale nelle attività dei servizi di alloggio e ristorazione con il 54,8 per cento, mentre per l'amministrazione pubblica e difesa, assicurazione sociale obbligatoria registra il 61,6 per cento (Istat, 2014).

12.4 Caso di studio e ArchIMEDe

Dopo una panoramica generale del fenomeno, si procede all'analisi dell'occupazione dei maschi e delle femmine per livello di istruzione, fascia d'età e condizione lavorativa ricorrendo ai dati ArchIMEDe.

Come nel resto del Paese, in regione Lombardia, l'occupazione femminile (43,1 per cento) è inferiore a quella maschile (56,9 per cento) soprattutto tra coloro che hanno l'istruzione primaria e un lavoro a tempo indeterminato, e l'istruzione primaria e secondaria e un lavoro autonomo (Figura 12.1). Più in particolare, per la

¹ I più giovani sono stati esclusi in quanto l'occupazione varia notevolmente secondo il livello di istruzione conseguito ed è più probabile che i più giovani stiano ancora studiando, in particolare nel ciclo terziario di istruzione.

classe d'età 25-34 anni, si osserva una quota inferiore di occupazione femminile tra coloro che hanno l'istruzione primaria e un lavoro a tempo indeterminato e l'istruzione secondaria per gli autonomi, mentre, al contrario, tra coloro che hanno una istruzione universitaria e un contratto a tempo determinato prevale l'occupazione femminile (Figura 12.2). Un andamento analogo si osserva nelle classi di età 35-44 e 45-54 anni (Figure 12.3 e 12.4).

Nella classe d'età 55-64 anni si osserva in generale un livello più basso di occupazione con un forte divario nel lavoro a tempo indeterminato tra coloro che hanno l'istruzione primaria e nel lavoro autonomo tra coloro che possiedono l'istruzione secondaria e universitaria; per questa classe d'età, invece, non si notano particolari differenze nel lavoro a tempo determinato (Figura 12.5²).

Per quanto riguarda i livelli di reddito individuale, si può notare considerando le diverse tipologie di contratto (Figure 12.6-12.8):

- che il lavoro a tempo indeterminato è più retributivo per gli uomini rispetto alle donne, maggiormente nella classe di età 55-64 anni e con istruzione universitaria. Il divario tende a diminuire tra coloro che hanno l'istruzione primaria e ad aumentare tra i più istruiti; inoltre, il divario è più basso tra i 25-34enni con istruzione universitaria e aumenta in maniera più sensibile nelle fasce d'età successive;
- il lavoro autonomo è più retributivo per gli uomini per ogni fascia d'età e livello di istruzione; il divario di genere è maggiore nella classe di età 45-54 anni più istruita (e con retribuzioni maggiori);
- Il lavoro a tempo determinato è più retributivo per gli uomini e, in particolare, tra coloro che hanno istruzione universitaria, registrando il divario minimo nella classe d'età più giovane e massimo in quella 55-64 anni. Anche per l'istruzione primaria e secondaria la differenza maggiore si osserva tra gli over 55.

Dalle *box plot map* del reddito medio per comune, si può osservare per la classe di età 25-34 anni una maggiore concentrazione di aree con divario elevato tra maschi e femmine nella parte Nord e Nord Est della regione Lombardia (province di Sondrio, Bergamo e Brescia) e nella provincia di Pavia (Figura 12.9). All'aumentare dell'età, il divario tende ad assumere valori più bassi nella provincia di Sondrio con una maggiore concentrazione nell'arco mediano: viene dunque coinvolta anche la parte Nord occidentale della regione con valori molto elevati nella provincia di Milano (Figure 12.10-12.12). Le *Lisa cluster map* individuano differenti *cluster*. I *cluster* di tipo *High-High* con valori molto elevati del fenomeno e forte similarità sono presenti nella parte Est della Lombardia per la classe d'età 25-34 anni, in particolare la provincia di Brescia si caratterizza per comuni con elevati divari e autocorrelazione positiva. I comuni a Nord della provincia di Brescia con occupati di 55-64 anni si caratterizzano per bassi valori del fenomeno e autocorrelazione positiva (*Low-Low*) nei comuni limitrofi. Il comune di Milano presenta alti valori di divario di reddito individuale diversamente dai comuni limitrofi (*High-Low*) nella fascia di età 25-34 anni (Figura 12.13). Al crescere dell'età, in particolare nella classe d'età 45-54 anni, i comuni del milanese e della Brianza tendono a valori molto alti di divario (Figure 12.14-12.15). Ciò invece non accade per l'ultima classe di età dove i valori molto alti dell'indicatore e autocorrelazione

² Nelle figure 12.1, 12.2, 12.3, 12.4 e 12.5 la somma dei valori percentuali non dà 100 perché non sono state prese in considerazione tutte le condizioni lavorative.

positiva si registrano nella parte inferiore delle province di Como e Lecco (Figura 12.16). Secondo i risultati del Censimento delle industrie e dei servizi 2011, sono proprio le province di Milano e Monza Brianza a registrare una maggiore frequenza di figure direttive e di impiegati per l'accentuata presenza di realtà produttive, in genere società di capitali. Nel resto del territorio lombardo, dove è maggiore la concentrazione di operai, si osserva una minore differenza di reddito complessivo tra maschio e femmina.

La popolazione lombarda di sesso femminile di età 25-54 anni è attorno alle 1,8milioni di unità; il 69,0 per cento di essa risulta occupata. La tipologia familiare prevalente in cui vivono le donne è costituita dalle coppie con figli (49,8 per cento) seguite dalle unipersonali (13,1 per cento). Il 76 per cento delle donne che hanno istruzione universitaria lavora, contro il 75,4 e il 60,0 per cento di quelle con istruzione rispettivamente secondaria e primaria. Le donne con figli hanno un più basso livello di occupazione. In particolare, hanno minori possibilità di accesso al mondo del lavoro le donne che vivono in coppia con figli (35,0 per cento) e le monogenitore (28,0 per cento), mentre risultano più avvantaggiate le donne senza figli che vivono in coppia (26,0 per cento) e le unipersonali (24,0 per cento) (Figura 12.17).

Per quanto riguarda l'analisi per quartili di reddito e confrontando le diverse tipologie familiari (Figura 12.18), si nota che le donne che vivono in coppia con figli presentano la percentuale più elevata (28,2 per cento) nel primo quartile - meno ricco - mentre le donne che vivono da sole e quelle monogenitore presentano la percentuale più elevata nell'ultimo quartile - più ricco (rispettivamente 37,0 e 28,4 per cento).

In riferimento alla classe di età, sono le donne di 45-54 anni che con maggiore frequenza hanno redditi alti (29,7 per cento appartengono all'ultimo quartile), mentre appartengono più frequentemente al primo quartile le più giovani 25-34enni (27,7 per cento).

Infine, nell'ultimo quartile si collocano anche le donne con istruzione universitaria (43,0 per cento) e, al contrario, le donne con istruzione primaria si collocano maggiormente nel primo quartile (32,9 per cento).

La quota di donne con figli piccoli che lavora è pari al 63,8 per cento. È interessante sottolineare che la presenza di figli di 0-14 anni non sembra incidere negativamente sull'occupazione femminile. Infatti, tra le donne occupate le differenze tra coloro che hanno figli di 0-14 anni e quelle con figli grandi sono inferiori ai 2 punti percentuali: la quota di donne che lavora e che ha figli piccoli è pari al 48,8 per cento, mentre il 51,2 per cento lavora e ha figli grandi.

12.5 Conclusioni

In generale si osserva che il divario di occupazione tra maschi e femmine tende a diminuire all'aumentare del livello di istruzione e nelle età più giovani. Per i contratti a tempo indeterminato il divario di genere a livello di occupazione è superiore tra coloro che hanno l'istruzione primaria, ed è quasi nullo per l'istruzione secondaria e universitaria. Le donne, però, per conciliare lavoro e famiglia spesso cercano un impiego in quei settori della pubblica amministrazione che richiedono un diploma (non necessariamente una laurea) e che garantiscono orari e condizioni lavorative più flessibili. La scelta di proseguire gli studi, infatti, sembra connessa all'importanza che si attribuisce all'autonomia economica femminile e alla possibi-

lità di poter conciliare lavoro e famiglia. A livello di reddito, il divario tra i sessi³ risulta più evidente tra coloro che hanno un'età e un'istruzione superiore. Osservando le donne, emerge che esse risultano prevalentemente occupate, con vantaggio, anche per il reddito, delle donne che vivono in coppia senza figli e delle single. Tra le donne con figli, è interessante notare che le monogenitore partecipano maggiormente al mercato del lavoro e percepiscono più frequentemente redditi più alti rispetto alle donne con figli che vivono in coppia; probabilmente, le madri che vivono da sole, non potendo contare su un altro reddito, hanno maggiore necessità a lavorare. La presenza di figli di 0-14 anni non sembra poi costituire un deterrente all'accesso al mercato del lavoro rispetto alle donne con figli più grandi. L'utilizzo dei dati amministrativi ha, dunque, costituito un valido strumento per approfondire a livello locale un importante fenomeno: quello delle differenze di genere. Pertanto, pur con le opportune cautele, l'utilizzo degli archivi amministrativi consente ai *policy maker* di poter conoscere meglio la realtà territoriale di appartenenza e prendere delle decisioni coerenti e appropriate ai bisogni locali.

Figura 12.1 - Occupati 25-64enni, per tipologia di occupazione, sesso e livello di istruzione. Lombardia, 2012 (valori %)

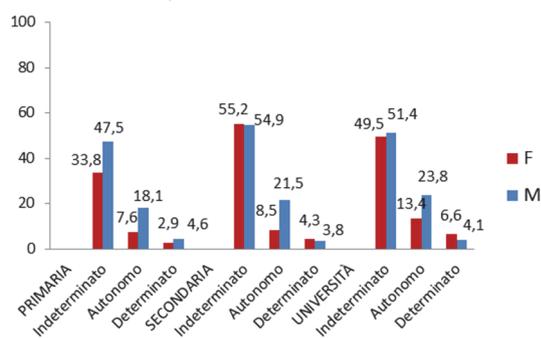


Figura 12.2 - Occupati 25-34enni, per tipologia di occupazione, sesso e livello di istruzione. Lombardia, 2012 (valori %)

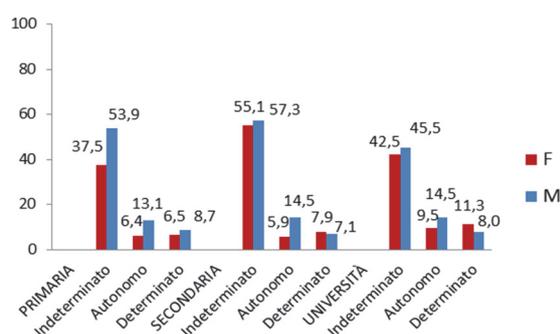


Figura 12.3 - Occupati 35-44enni, per tipologia di occupazione, sesso e livello di istruzione. Lombardia, 2012 (valori %)

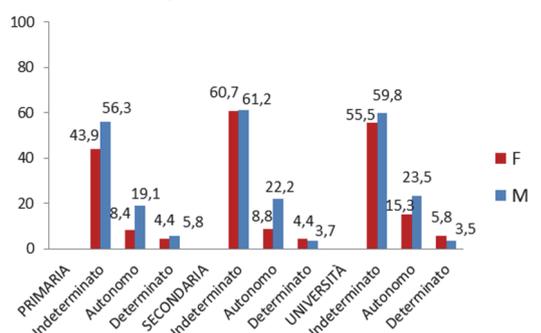
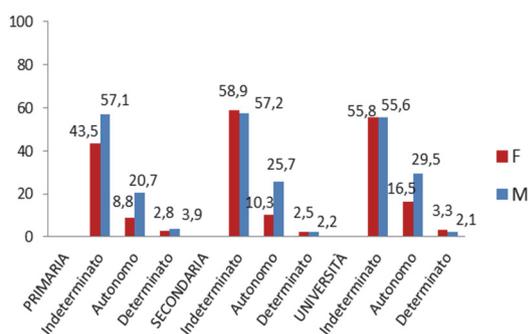


Figura 12.4 - Occupati 45-54enni, per tipologia di occupazione, sesso e livello di istruzione. Lombardia, 2012 (valori %)



³ Questa differenziazione di reddito complessivo tra maschi e femmine non sempre denota una condizione di fragilità effettiva per le donne, giacché in tali circostanze entrambi i sessi potrebbero presentare redditi elevati.

Figura 12.5 - Occupati 55-64enni, per tipologia di occupazione, sesso e livello di istruzione. Lombardia, 2012 (valori %)

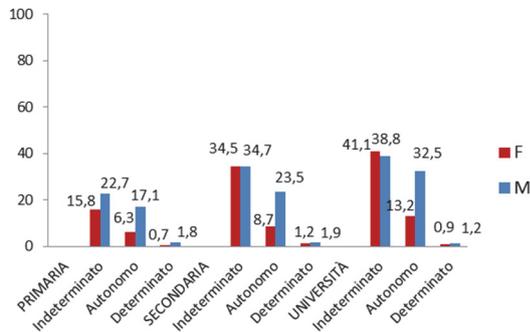


Figura 12.6 - Reddito individuale dei contratti a tempo indeterminato per sesso, titolo di studio ed età. Lombardia, 2012

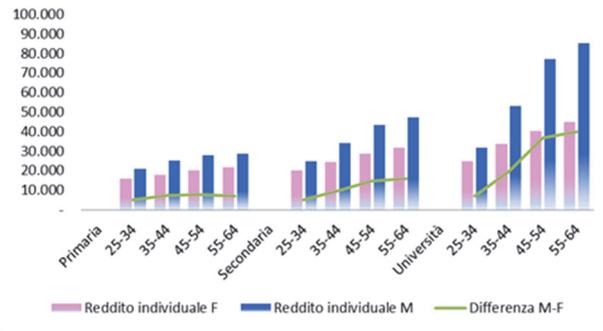


Figura 12.7 - Reddito individuale degli autonomi per sesso, titolo di studio e fascia d'età. Lombardia, 2012

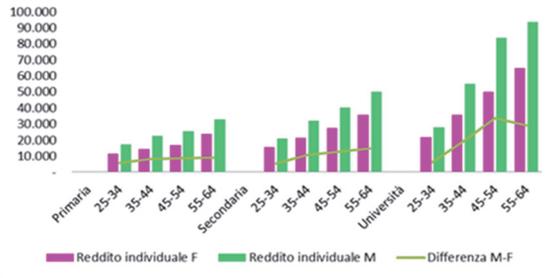


Figura 12.8 - Reddito individuale dei contratti a tempo determinato per sesso, titolo di studio e fascia d'età. Lombardia, 2012

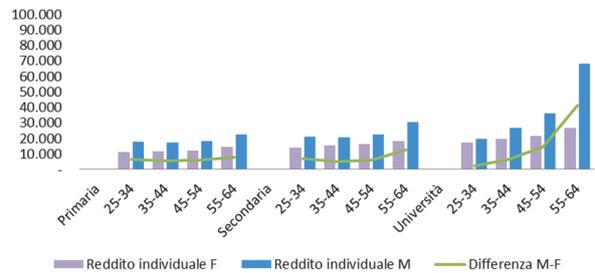


Figura 12.9 - Divario maschi e femmine 25-34enni per reddito individuale a livello comunale. Lombardia, 2012 (Box Plot Map)

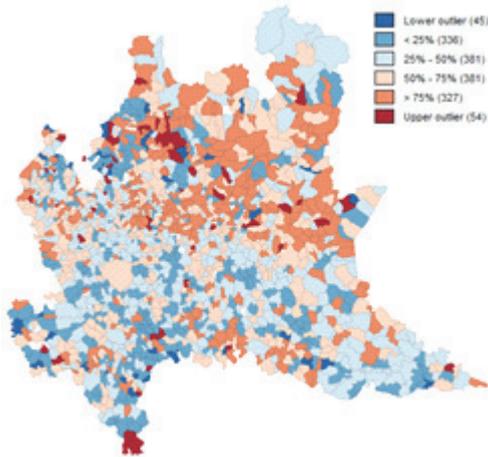


Figura 12.10 - Divario maschi e femmine 35-44enni per reddito individuale a livello comunale. Lombardia, 2012 (Box Plot Map)

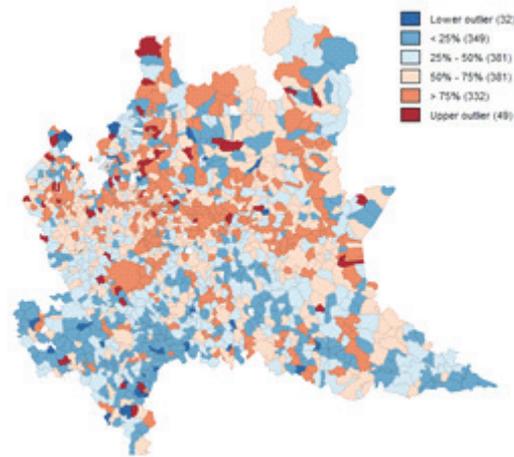


Figura 12.11 - Divario maschi e femmine 45-54enni per reddito individuale a livello comunale. Lombardia, 2012 (Box Plot Map)

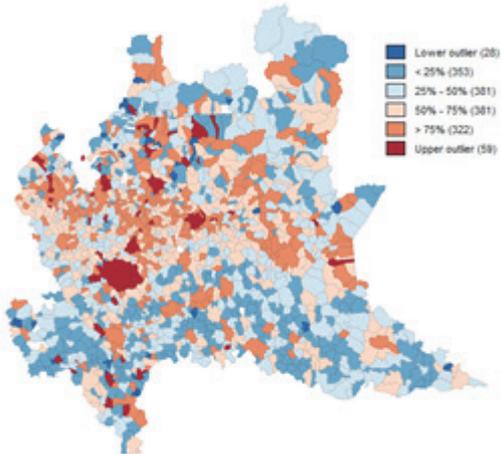


Figura 12.12 - Divario maschi e femmine 55-64enni per reddito individuale a livello comunale. Lombardia, 2012 (Box Plot Map)

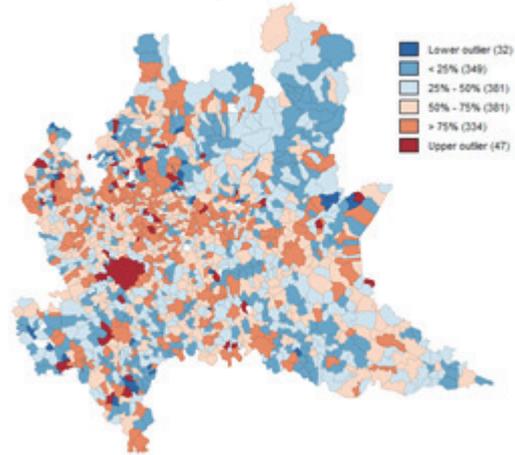


Figura 12.13 - Divario maschi e femmine di 25-34 anni per reddito individuale a livello comunale. Lombardia, 2012 (LISA Cluster Map)

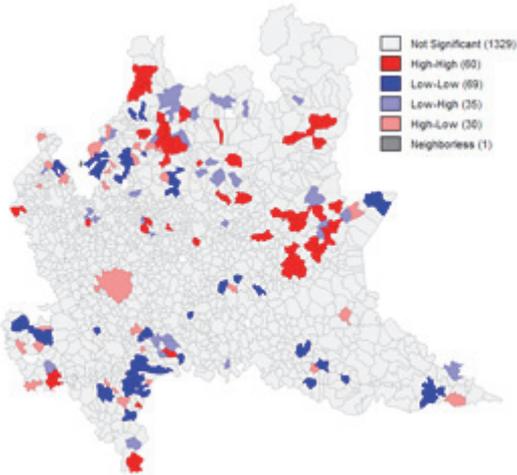


Figura 12.14 - Divario maschi e femmine di 35-44 anni per reddito individuale a livello comunale. Lombardia, 2012 (LISA Cluster Map)

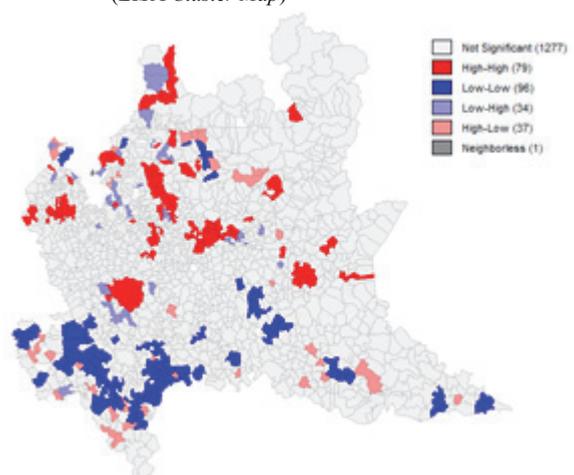


Figura 12.15 - Divario maschi e femmine di 45-54 anni per reddito individuale a livello comunale. Lombardia, 2012 (Lisa Cluster Map)

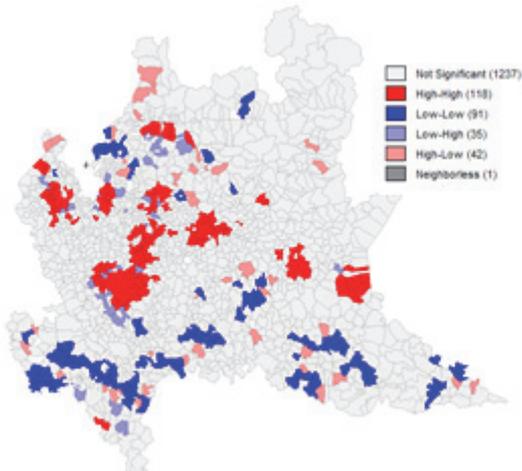


Figura 12.16 - Divario maschi e femmine di 55-64 anni per reddito individuale a livello comunale. Lombardia, 2012 (Lisa Cluster Map)

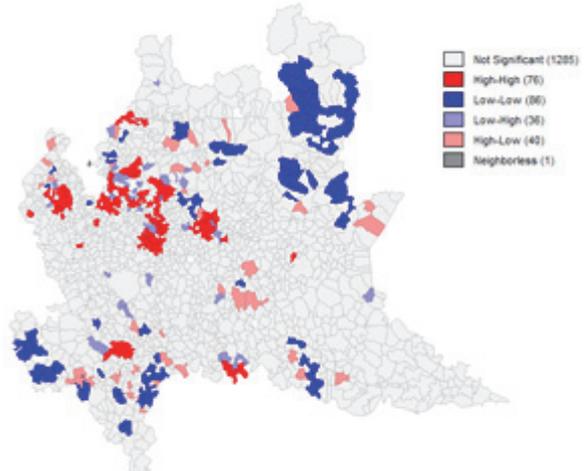


Figura 12.17 - Donne occupate e non per titolo di studio, per tipologia familiare e per classi di età. Lombardia, 2012 (composizioni %)

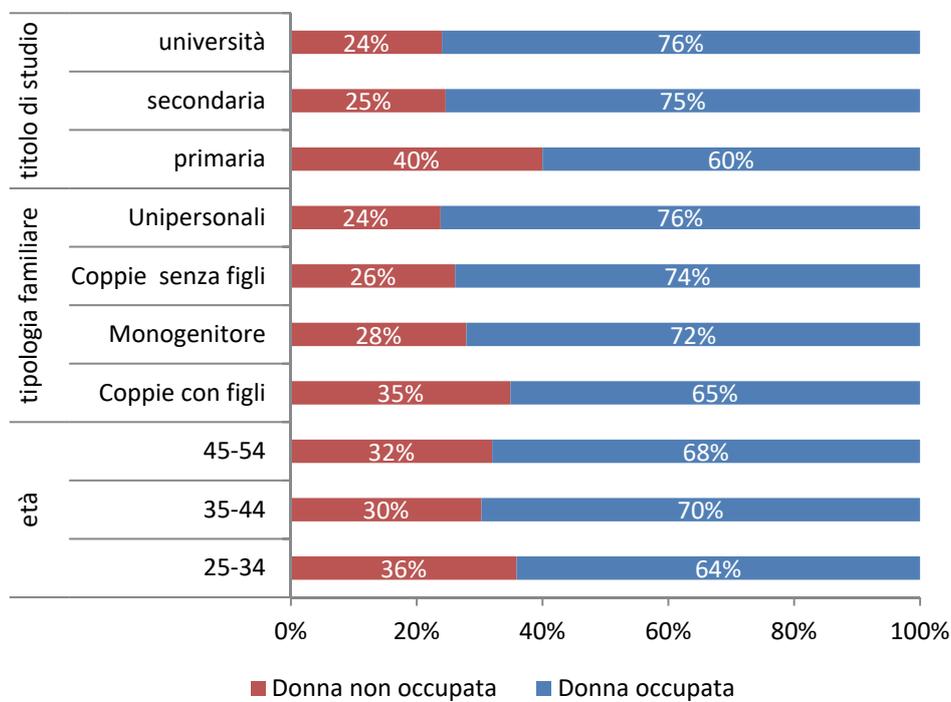
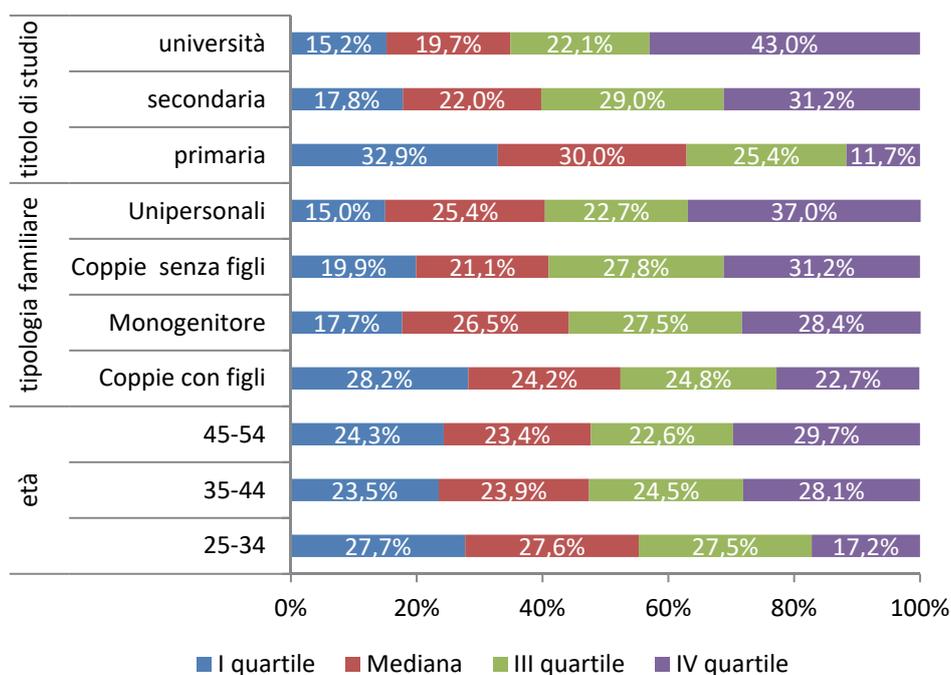


Figura 12.18 - Distribuzione percentuale delle donne nei diversi quartili di reddito per titolo di studio, per classe di età e per tipologia familiare. Lombardia, 2012 (composizioni %)



CAPITOLO XIII

NEET E QUASI-NEET IN LOMBARDIA¹

di Simona Ballabio, Arianna Carra, Maria Elena Comune, Daniela Ferrazza,
Flavio Verrecchia, Alberto Vitalini, Lorena Carmen Maria Viviano.

13.1 Introduzione

Rispetto alle generazioni precedenti, la transizione scuola-lavoro della *Generazione del millennio* è più incerta e – comunque – rallentata; pertanto questa fase tende a gravare maggiormente sui percorsi di vita dei giovani. L'ampliarsi dello stato di indeterminatezza, la situazione sospesa e non definita caratterizzata dall'inattività comporta, infatti, una deprivazione socio-economica difficilmente estinguibile nel breve periodo.

In questo contesto l'osservazione della struttura e della dinamica dei Neet rappresenta un opportuno tentativo di misura della vulnerabilità delle nuove generazioni anche in termini di mancate opportunità di inserimento e partecipazione al sistema scuola-lavoro e di *welfare* più in prospettiva.

Negli ultimi anni i Neet da categoria statistica residuale (Gracey S. e Kelly S., 2010) si sono trasformati in un topic autonomo, che orienta risorse intellettuali e politiche di diverso tipo. I principali studi pubblicati a riguardo (Eurofund, 2012, 2016) e le più autorevoli statistiche diffuse (p.es. quelle pubblicate da Eurostat o da Istat) restituiscono un quadro informativo approfondito, con un elevato standard qualitativo e, in molti casi, confrontabile a livello internazionale. Tuttavia, sono stime provenienti da indagini campionarie rappresentative a livello di macro-aree territoriali. Queste ultime non permettono di scendere a livelli di dettaglio ulteriori rispetto a quelli considerati nella progettazione delle indagini e questo ne indebolisce la valenza conoscitiva quando si passa dal piano teorico a quello della programmazione e progettazione delle politiche locali.

Grazie all'utilizzo di una fonte che si basa su dati amministrativi è possibile superare questo limite e rispondere a domande quali: come è articolata la popolazione dei Neet nelle singole unità amministrative? Quali specificità si rilevano nella distribuzione a livello locale?

L'obiettivo del presente contributo è proprio quello di dare risposta a queste domande utilizzando i dati del progetto ArchIMEDe. In particolare, si presenterà una sperimentazione concreta per lo studio del fenomeno e per il supporto alle politiche locali, illustrando la distribuzione territoriale dell'incidenza dei Neet in Lombardia nel 2013.

Non solo, dal *linkage* deterministico di basi dati amministrative si ottiene la disponibilità contemporanea di molteplici dimensioni relative ai soggetti di studio, tanto da consentire riflessioni sulle definizioni operative e sulle misure dei fenomeni oggetto di interesse che vengono comunemente adottate. Così, sfruttando le potenzialità euristiche dell'archivio, si è giunti anche a considerare il reddito come ulteriore dimensione, oltre a quelle tradizionalmente considerate, in grado di seg-

¹ Nel capitolo si riprendono solo alcuni aspetti inerenti alle sperimentazioni dei ricercatori di Milano sul tema dei Neet. Altri risultati sono presenti trasversalmente nelle parti introduttive del volume, in quanto le esemplificazioni sono state condotte proprio sul fenomeno dei Neet nel contesto lombardo (Cfr. Capitoli 3 e 6).

mentare la stessa popolazione dei Neet e di far luce sulle aree grigie a loro contorno, in modo da giungere a delineare un'ulteriore categoria concettuale definibile come Quasi-Neet (Ballabio *et al.*, 2017). In questo senso si presenteranno alcune caratteristiche più rilevanti anche di questa sub-popolazione che, specie nel nuovo contesto normativo che regola il mercato del lavoro, sembra essere emergente.

13.2. Neet: la tradizione

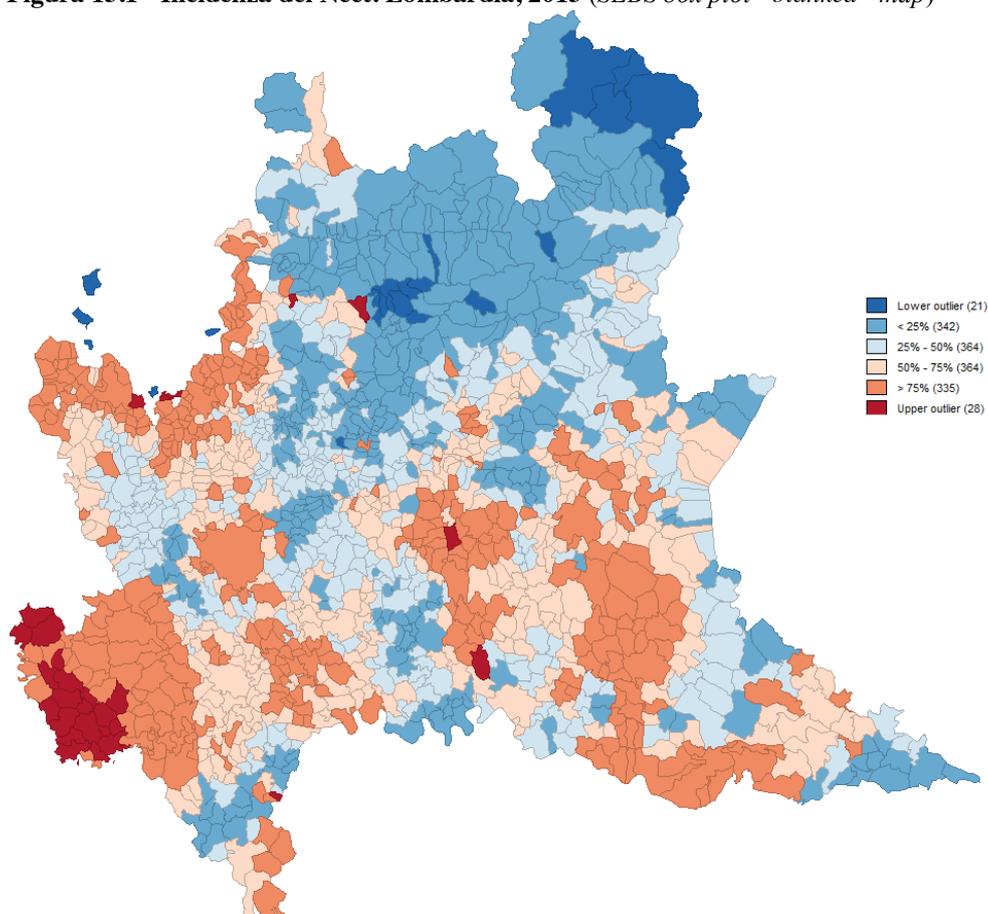
Partendo dalla definizione più comunemente adottata del fenomeno, che definisce i Neet come “giovani di età compresa tra i 15 e i 29 anni che non studiano e non lavorano”, si può osservare che per ciascuna dimensione considerata vi sia una corrispettiva e analoga fonte di informazioni in ArchIMEDe:

- “*i Neet sono giovani di età compresa tra i 15 e i 29 anni*”. Fonte: Lista anagrafica comunale (LAC) che riporta le caratteristiche anagrafiche di tutti i componenti dei nuclei familiari residenti, nonché loro dimensione e tipologia e da cui si può ricavare la sub-popolazione di giovani di interesse;
- “*che non studiano*”. Fonte: anagrafe degli studenti Miur che consentono di isolare e, dunque, togliere dal conteggio coloro che risultano iscritti alla scuola secondaria o all'università. Da sottolineare che, ad oggi, ArchIMEDe non include i dati della formazione professionale regionale;
- “*e non lavorano.*”. Fonte: Asia-Occupazione, posizioni Assicurative Ex Inpdap, lavoratori Autonomi in Agricoltura e lavoratori Domestici che consentono di individuare e, dunque, togliere dal conteggio coloro che mostrano segnali di attività sul mercato del lavoro.

Per la stima dei Neet da fonti secondarie, pertanto, alla stregua del blocco di marmo inerte da cui lo scultore fa emergere poco alla volta solo la forma che gli interessa, si delinea un percorso di filtri progressivi applicati alla base dati iniziale che fa emergere la *popolazione* di interesse.

Come si evince dalla cartografia (Figura 13.1)² – al netto della zona insubrica che a causa del lavoro oltre frontiera non è stata considerata (Cfr. Capitolo 3) – nel contesto lombardo si osserva una distribuzione non omogenea delle incidenze dei Neet a livello comunale. In particolare, si rilevano incidenze mediamente più contenute nell'area valtellinese. Mentre si osservano incidenze più elevate in quattro aree distinte: nella zona della Lomellina, nel basso bresciano e mantovano e nella zona al confine tra la provincia di Lodi e quella di Pavia. A queste si aggiungono altri due poli urbani di concentrazione del fenomeno: l'area milanese e quella bergamasca.

² La figura 13.1 corrisponde alla figura 6.7 del Volume. Pertanto per la descrizione e la corretta interpretazione di questo tipo di cartografia si rimanda al paragrafo 6.5.

Figura 13.1 - Incidenza dei Neet. Lombardia, 2013 (SEBS box plot “blanked” map)

Note: Cfr. Figura 6.7.

13.3 Una nuova categoria concettuale: i Quasi-Neet

Grazie alle sperimentazioni che si sono susseguite nel tempo è possibile mettere in luce in maniera pragmatica e sostanziale il contributo conoscitivo potenzialmente presente negli archivi amministrativi che, se debitamente trattati, possono rappresentare un valore aggiunto con riferimento non solo all’aumento del dettaglio territoriale, non solo all’accrescimento della portata informativa della produzione statistica, ma anche alla flessibilità nella “costruzione del dato” che non deve essere necessariamente pianificata ex-ante come avviene invece nelle indagini statistiche.

In questo senso, ad integrazione della tradizionale definizione di Neet, la ricchezza e l’articolazione dei dati d’archivio del progetto ArchIMEDe consentono di considerare ulteriori dimensioni che gettano una nuova luce sull’osservazione delle caratteristiche e delle dinamiche sociali della fascia dei giovani considerata (15-29enni). Ciò consente di acquisire ulteriori elementi informativi sia sui profili dei giovani oggetto di riflessione sia sul consolidarsi di nuovi paradigmi del mercato del lavoro in generale (i.e. microlavoro) che, accanto ai Neet, possono essere utilmente considerati per la progettazione di adeguati strumenti di intervento.

Nello specifico l’inserimento della variabile del reddito individuale, in combinazione con le altre dimensioni che delineano il fenomeno dei Neet, permette di individuare la nuova categoria concettuale dei Quasi-Neet. Quest’ultima si pone “a

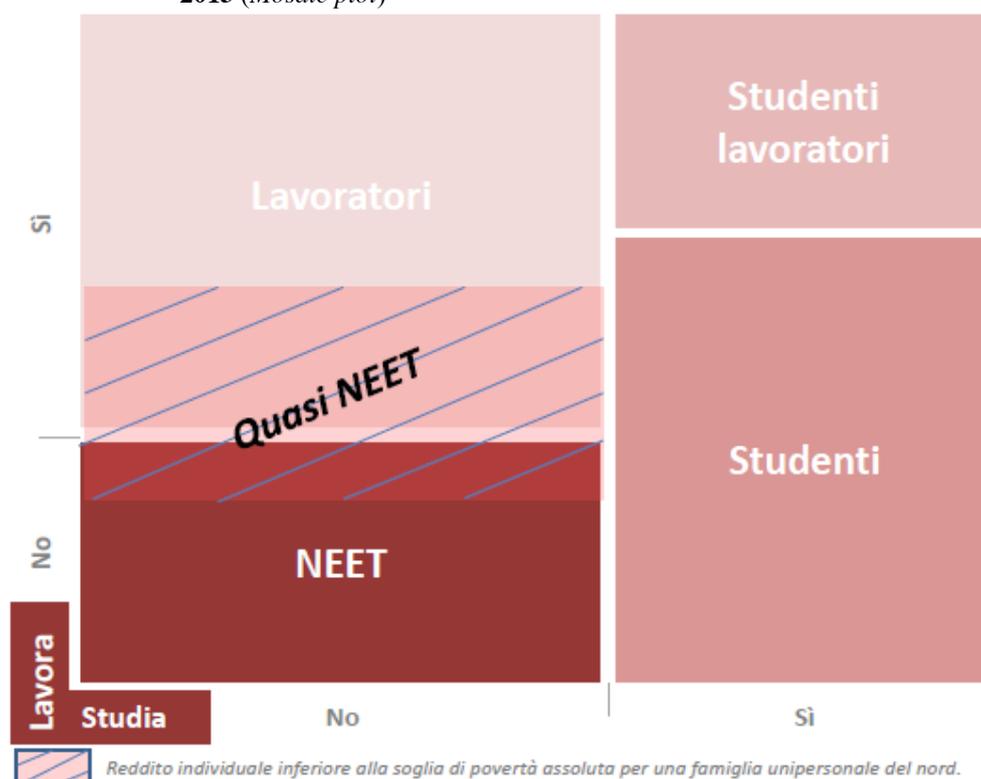
cavallo” delle categorie (ex)Neet e giovani Lavoratori (Figura 13.2) e include sia il sottogruppo degli (ex)Neet che presentano segnali di reddito individuale basso, sia quello dei giovani Lavoratori con un reddito individuale che non è in grado di garantire l’indipendenza economica³.

La considerazione della dimensione del reddito, inoltre, migliora la capacità di cogliere l’andamento nel tempo del fenomeno, poiché si tratta di una dimensione più reattiva rispetto ad altre (maggiormente legate alla normativa contrattuale) che rilevano i cambiamenti con un certo grado di differimento temporale.

L’emergente categoria dei Quasi-Neet è in possibile espansione e sembra essere particolarmente connessa alle nuove dinamiche del mercato del lavoro e ai nuovi orizzonti occupazionali di lavoro accessorio o intermittente.

In questo senso, per la progettazione di adeguati strumenti di intervento si delinea una ulteriore popolazione target (rispetto a quella dei Neet) caratterizzata, sebbene a diverso titolo, da un potenziale elevato grado di vulnerabilità sociale e materiale.

Figura 13.2 - Giovani 15-29enni: Lavoro, studio e reddito individuale. Lombardia, 2013 (Mosaic plot)



Note: Cfr. Ballabio et al. (2017).

A fini descrittivi, nei grafici sotto riportati (Figura 13.3) le categorie di Neet, Quasi-Neet e resto della popolazione 15-29enne sono messe in relazione ad alcune variabili socio-demografiche standard, quali: relazioni di parentela, cittadinanza, genere (dei soli residenti di cittadinanza straniera), titolo di studio, età.

³ Più specificatamente è stata considerata la soglia di povertà assoluta per una famiglia unipersonale del Nord Italia (Istat, 2014).

Inoltre, per considerare gli altri pilastri su cui si fonda il *welfare* italiano – rappresentati dalla rete familiare e da quella socio-assistenziale – e al contempo segmentare rispetto al grado di vulnerabilità le tre sottopopolazioni si è anche considerato il reddito pro-capite familiare equivalente.

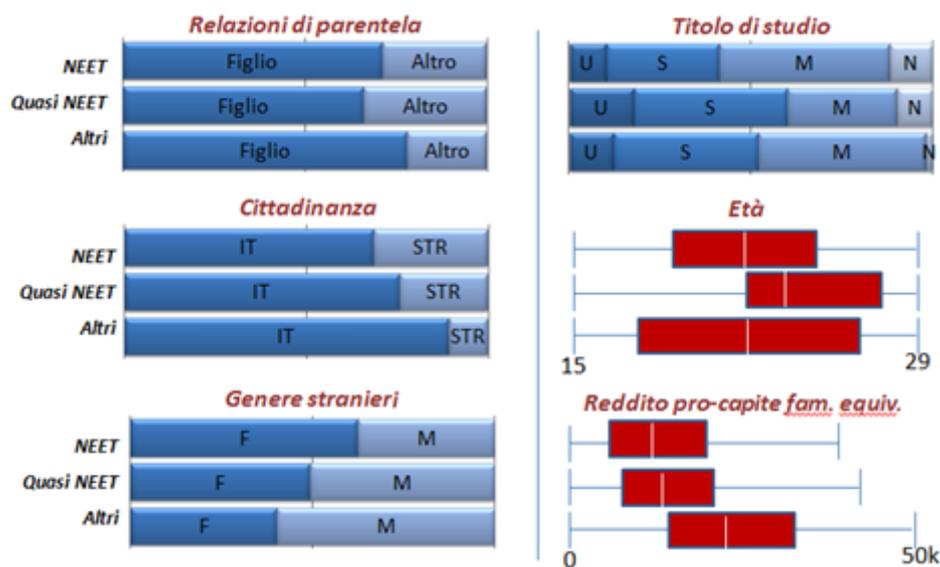
Grazie al reddito familiare equivalente (Figura 13.3 – in basso a destra) si può osservare come la categoria dei Quasi-Neet abbia una struttura omogenea – rispetto alla dimensione considerata – a quella dei Neet e diversa dalle famiglie degli altri giovani. Questo aspetto supporta la decisione di estendere gli studi sulla vulnerabilità giovanile, connessa alle difficoltà di inserimento nel sistema scuola-lavoro, anche al gruppo dei Quasi-Neet.

Per quanto riguarda le relazioni di parentela, in tutte le categorie la maggioranza dei giovani 15-29enni vive all'interno del nucleo familiare di origine. Anche in questo caso non sembra emergere una sostanziale differenza tra Neet e Quasi-Neet rispetto a questa caratteristica.

Differenze tra Neet e Quasi-Neet si osservano invece rispetto all'età e al titolo di studio: i Quasi-Neet infatti sono tendenzialmente più anziani e con un titolo di studio più elevato.

Se si considera la cittadinanza, i Quasi-Neet, rispetto ai Neet, sono in misura maggiore cittadini italiani. Concentrando poi l'attenzione sul genere dei giovani stranieri emerge che le ragazze sono esposte ad un rischio di vulnerabilità superiore rispetto ai loro coetanei.

Figura 13.3 - Distribuzione, per relazioni di parentela, per cittadinanza, per genere degli stranieri, per titolo di studio, per età e per reddito pro-capite familiare equivalente dei Neet, Quasi-Neet e Altri. Lombardia, 2013



Note: Cfr. Ballabio et al. (2017); **Relazioni di parentela** - Figlio: giovane che vive con almeno uno dei genitori; altro: altra tipologia di relazione con l'intestatario della scheda anagrafica; **Titolo** - U: Universitario; S: Superiore; M: Medie e elementari; N: Nessun titolo; **Genere stranieri** - F: Femmine; M: Maschi; **Cittadinanza** - IT: Italiani; STR: Stranieri; **Categorie** - Neet: Neet senza reddito; Altri: 15-29enni non Neet e non Quasi-Neet.

13.4 Conclusioni e sviluppi futuri

Il lavoro svolto si è basato sull'utilizzo a fini sperimentali dei dati del progetto ArchIMEDe.

Dall'esplorazione dei dati d'archivio sui residenti lombardi sono emerse alcune interessanti riflessioni, che suggeriscono di considerare con maggiore attenzione le definizioni operative e le misure dei fenomeni che vengono comunemente adottate. In uno scenario d'integrazione e/o di possibile transizione da fonti d'indagine a fonti d'archivio della statistica ufficiale italiana, il primo passo per lo studio dei fenomeni è infatti proprio la problematizzazione delle misure e delle definizioni operative esistenti.

Anche se i dati utilizzati – da fonte secondaria – non sono strutturalmente simili a quelli delle indagini campionarie dell'Istat e nemmeno direttamente confrontabili, le nuove frontiere della statistica ufficiale delineano una interrelazione sempre più fitta che impone una lettura dei fenomeni a partire da fonti diverse per trarre il massimo delle informazioni e della qualità possibile.

La riconsiderazione delle definizioni operative e delle misure comunemente adottate quali *proxy* della vulnerabilità giovanile è diventato quindi nel capitolo un obiettivo autonomo.

Nello specifico, sfruttando le potenzialità euristiche dell'archivio si è giunti a considerare il reddito come ulteriore dimensione, oltre a quelle tradizionalmente considerate, in grado di segmentare la popolazione dei Neet e di far luce sulle aree grigie a loro contorno, così da giungere a delineare ulteriori categorie concettuali come quella descritta in questo capitolo dei Quasi-Neet – giovani non impegnati in un percorso formativo e con un reddito individuale inferiore alla soglia di povertà assoluta per una famiglia unipersonale del nord (Cfr. Ballabio *et al.* 2017).

Nell'ambito della riconsiderazione delle definizioni operative, un ulteriore sviluppo – utilizzando le informazioni sul reddito – riguarda l'approfondimento di un sottogruppo specifico che potrebbe presentare elevati livelli di vulnerabilità in relazione non solo alla non appartenenza al sistema formazione-lavoro ma anche rispetto all'assenza di una rete di protezione socio-parentale: i Nieet (*Not Income, in Employment, Education or Training*). Si tratta quindi di giovani 15-29enni che non solo non studiano o lavorano ma non possono nemmeno contare su altre fonti di reddito (Cfr. Ballabio *et al.*, 2018).

CAPITOLO XIV

NON PROFIT IN LOMBARDIA: UNA RISPOSTA RESILIENTE?¹

di Stefania Della Queva, Manuela Nicosia e Sabrina Stoppiello

14.1 Introduzione

Il 9° Censimento generale dell'industria e dei servizi e Censimento delle istituzioni non profit (2011) ha messo in luce il ruolo di primaria importanza che il settore non profit riveste in Lombardia, visto il peso che occupa in termini di unità locali, pari a 53.934 unità, quasi il 6 per cento di tutte le unità giuridico economiche che hanno sede nella Regione. In particolare, la Lombardia ha registrato in termini assoluti il numero maggiore di istituzioni non profit (INP) attive nel 2011 pari a 46.141 (15,3 per cento del totale nazionale) e di unità locali, pari al 15,5 per cento del totale nazionale. La consistenza del personale retribuito (dipendenti e lavoratori esterni) delle unità locali lombarde è di circa 210mila unità (pari al 22,1 per cento del totale nazionale) mentre i volontari costituiscono il 16,7 per cento del totale nazionale (quasi 800mila individui).

Lo studio si pone l'obiettivo di delineare il ruolo del settore non profit in Lombardia come risorsa in situazioni di vulnerabilità sociale ed economica, soprattutto in contesti territoriali disagiati. In particolare, attraverso un'analisi sperimentale², il lavoro ha analizzato la presenza e la diffusione territoriale di alcune tipologie di istituzioni non profit, alla luce di alcune forme di vulnerabilità familiare della popolazione residente nei sistemi locali del lavoro (Sll) lombardi.

14.2 Dati e metodo

I dati utilizzati, relativi alle INP e riferiti al 2011, derivano dal 9° Censimento generale dell'industria e dei servizi e Censimento delle istituzioni non profit. A questi si affiancano i dati amministrativi del progetto ArchIMEDe, riferiti all'anno 2012, sulle "Condizioni socio-economiche delle famiglie", che contiene dati elementari funzionali alla ricerca sociale ed economica, alla programmazione territoriale e settoriale, nonché alla valutazione delle politiche pubbliche³.

Nella prima fase del progetto, sono stati elaborati i principali indicatori utili a descrivere i due fenomeni. Per quanto riguarda il settore non profit sono stati elaborati indicatori relativi a:

- struttura e diffusione sul territorio⁴: dimensione delle INP in termini di volontari e lavoratori retribuiti, presenza del settore in termini di INP e volontari;

¹ Il capitolo è frutto della collaborazione degli autori. In particolare, i paragrafi 14.2, 14.3.2 sono da attribuirsi a Stefania della Queva, il paragrafo 14.3.1 è da attribuirsi a Manuela Nicosia.

² Data la natura sperimentale del lavoro per quanto riguarda gli Sll dell'area insubrica si rimanda alle considerazioni generali del Capitolo 3.

³ L'uso di dati di due anni di riferimento differenti si basa sull'ipotesi di lentezza dei cambiamenti strutturali.

⁴ Percentuale di istituzioni non profit INP con 20 volontari e più; percentuale di INP con 20 lavoratori retribuiti e più, unità locali delle INP per km quadrato; numero di volontari ogni 1000 abitanti; percentuale di INP nate prima del 1990; percentuale di INP nate dopo il 2000.

- anno di costituzione delle INP;
- settore di attività prevalente⁵;
- tipo di attività economica⁶;
- tipologia di disagio dei destinatari⁷;
- *mission* o finalità⁸ delle attività realizzate.

Per quanto riguarda la condizione di vulnerabilità familiare sono stati elaborati indicatori relativi a:

- condizione del capofamiglia⁹;
- struttura familiare¹⁰.
- caratteristiche dei componenti¹¹
- condizione economica¹².

Nella seconda fase, è stata condotta un'analisi in componenti principali (ACP), distinguendo l'area relativa agli indicatori sulle istituzioni non profit da quella relativa agli indicatori di vulnerabilità familiare. La lente dei Sistemi Locali del Lavoro, definiti dall'Istat secondo i dati del Censimento della popolazione e delle abitazioni 2011 (Istat, 2015a), è stata utilizzata per l'aggregazione dei dati. In particolare, sono stati selezionati i 51 SL "puri" i cui comuni rientrano tutti nei confini della regione Lombardia e non sono stati considerati i 6 SL a cavallo con i confini regionali¹³.

L'analisi ha l'obiettivo di individuare, da un lato, le aree nelle quali il non profit assume un ruolo di sostegno sociale delle comunità residenti, dall'altro, quelle nelle quali le famiglie vivono situazioni di disagio lavorativo, sociale ed economico per verificare la funzione svolta dal settore non profit in termini di resilienza alla vulnerabilità familiare.

Gli strumenti di *data visualization*, quali le *box plot map* e le *cluster map*, hanno consentito di rappresentare la distribuzione territoriale di questi fenomeni e di mostrare l'esistenza di eventuali associazioni spaziali.

14.3 Risultati

14.3.1 I SL del non profit

Il primo indice sintetico, costruito sulle variabili descrittive del non profit, indica la presenza del "*settore non profit per il disagio*". Delinea, infatti, istituzioni che pongono attenzione alla popolazione con forme differenziate di disagio, da quello legato alla malattia, che riguarda tutta la popolazione, indipendentemente

⁵ Percentuale di INP attive nei settori: Cultura, sport e ricreazione; Sanità; Assistenza sociale; Sviluppo economico e coesione sociale; Filantropia e promozione del volontariato; Religione.

⁶ Percentuale di INP market (INP che sostengono i propri costi attraverso entrate derivanti da contratti e/o convenzioni con le istituzioni pubbliche e ricavi derivanti da vendita di beni e servizi); percentuale di INP di pubblica utilità; percentuale di INP con finanziamento prevalente pubblico.

⁷ Percentuale di INP che si occupano delle seguenti categorie di disagio: devianza, detenzione carceraria, dipendenze, disabilità e non autosufficienza, malattia, immigrazione e nomadismo, povertà ed esclusione sociale, abusi e molestie, prostituzione, disagio psico-sociale.

⁸ Percentuale di INP le cui attività sono state realizzate per la promozione e la tutela dei diritti; percentuale di INP le cui attività sono state realizzate per il sostegno ed il supporto di soggetti deboli e/o in difficoltà.

⁹ Percentuale di famiglie con capofamiglia che non ha lavorato almeno un giorno nell'anno e non ha reddito da pensione; percentuale di famiglie con capofamiglia con nessun titolo di studio o la licenza elementare.

¹⁰ Percentuale di famiglie con almeno un figlio minore dai 0 ai 14 anni; percentuale di famiglie con almeno una persona anziana di 65 anni e oltre; percentuale di famiglie numerose con più di 5 componenti.

¹¹ Percentuale di famiglie con almeno un componente con cittadinanza straniera o apolide; percentuale di famiglie con almeno un componente tra i 18 e i 39 anni che non ha reddito né studia; percentuale di coppie con figli nelle quali un solo componente è percettore di reddito.

¹² Reddito medio equivalente familiare pro-capite (con polarità invertita).

¹³ L'analisi include i SL dei Comuni Transfrontalieri dell'area insubrica (confinanti con la Svizzera).

dalla struttura demografica e dalla presenza di categorie specifiche, ad altri tipi di disagio, relativi a particolari categorie della popolazione, quali gli immigrati e i detenuti. L'indice descrive infatti, da un lato, si individuano le istituzioni non profit nate prima degli anni '90, la cui attività si basa sull'apporto di un numero consistente di volontari, di orientamento solidaristico, con attività indirizzate alla malattia come categoria di disagio, oppure legate al soddisfacimento di fini culturali-ricreativi. Dall'altro, istituzioni di nuova generazione (costituite dopo il 2000), dedite a categorie specifiche di disagio (povertà, abusi e molestie, immigrazione e detenzione), dove tra le finalità spicca la promozione e tutela dei diritti.

La lettura del territorio lombardo attraverso questa sintesi rivela una diversa caratterizzazione dei sistemi locali rispetto al tipo di non profit attivo (Figura 14.1). Quelli situati sul versante montano ad est (Livigno, Edolo, Tirano, Clusone) e ad ovest (Chiavenna, Morbegno, Menaggio) presentano infatti una forte presenza del volontariato storico, che svolge più funzioni sul territorio. Spicca il sistema locale di Sondalo, dove il 40 per cento delle INP conta su un consistente numero di volontari (più di 20), il 70 per cento delle INP dedite al disagio si occupa di malati, così come quelli di Ponte di Legno e Porlezza e Limone sul Garda (Sistemi locali con valore dell'indicatore del primo quartile). Diversamente si caratterizza tutta la fascia dei sistemi locali che aggregano i comuni delle aree urbane di Como, Milano, Pavia, Vigevano, Lodi, Cremona, Brescia, Mantova, connotati appunto da un territorio più urbanizzato, dove diventa più urgente la tutela dei diritti e l'attenzione a specifici disagi (Sistemi locali con valore dell'indicatore dell'ultimo quartile). Spiccano in questo contesto Asola dove, tra le INP dedite al disagio, il 43 per cento si occupa di povertà, e il sistema locale di Milano dove le stesse rappresentano il 30 per cento (Sistemi locali con valore *outlier*, verso l'alto, dell'indicatore).

L'analisi dell'autocorrelazione spaziale (Figura 14.2) conferma la contrapposizione territoriale tra il sistema-montagna e il sistema-città¹⁴. Quest'ultimo emerge come l'area 'calda' dove si concentrano diverse tipologie di disagio cui si rivolgono le istituzioni non profit presenti sul territorio. Si evidenzia inoltre la diversità del sistema locale di Casalmaggiore rispetto ai comuni vicini (Sistema locale con valore basso dell'indicatore).

Figura 14.1 - “non profit tra ieri e oggi”, 2011
(Box plot map)

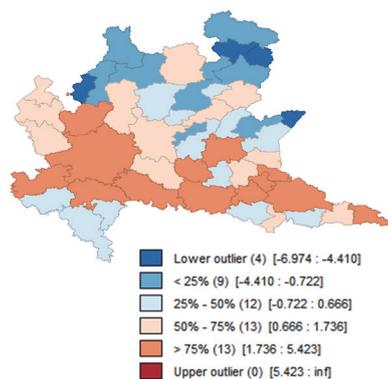
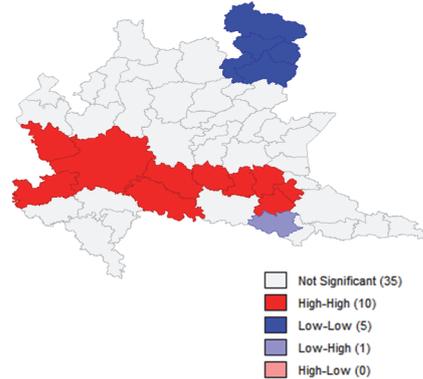


Figura 14.2 - “non profit tra ieri e oggi”, 2011
(LISA cluster map)



Il secondo indice sintetico rappresenta “*il non profit verso la specializzazione*”. Esso consente infatti di individuare, da un lato, la presenza del volontariato

¹⁴ N.B. Il valore alto dell'indicatore (in rosso nella cartina, “high” nella legenda) individua il polo positivo della componente, il valore basso (in blu nella cartina, “low” nella legenda) il polo negativo.

storico che si dedica al supporto e sostegno dei soggetti deboli, di orientamento solidaristico, basato sul finanziamento pubblico (Figura 14.3: Sistemi locali con valore dell'indicatore del primo quartile). Qui spiccano i sistemi locali di Vilminore di Scalve e Ponte di Legno, dove le INP solidaristiche rappresentano tra l'80 per cento e il 90 per cento del totale (la media per SI è 69,4 per cento), le INP con il tipo di *mission* citata sono il 57 per cento (la media regionale è 47 per cento). Dall'altro, meno diffuso perché più recente, il non profit che caratterizza alcuni sistemi locali con collocazione sparsa sul territorio, dove la presenza di INP che svolgono attività sanitaria supera la media dei SI (5,8 per cento) con punte del 10 per cento in sistemi locali come Sannazzaro De' Burgondi, Sermide e Viadana. In questi SI spiccano inoltre le INP con attività economica *market-oriented*¹⁵ e quelle dedite al disagio psico-sociale (Sistemi locali con valore dell'indicatore dell'ultimo quartile). L'analisi dell'autocorrelazione spaziale (Figura 14.4) mette in evidenza il sistema locale di Tirano (valore alto dell'indicatore vicino a SI con valore basso), dove le INP sanitarie sono il 7 per cento, e quelle dedite al disagio psico-sociale il 12 per cento (la media dei SI è 5 per cento).

Figura 14.3 - “non profit verso la specializzazione”, 2011 (Box plot map)

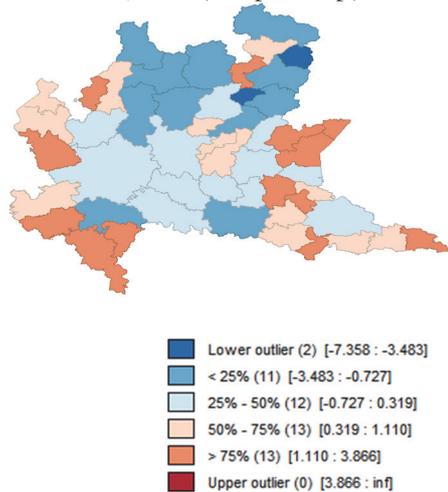
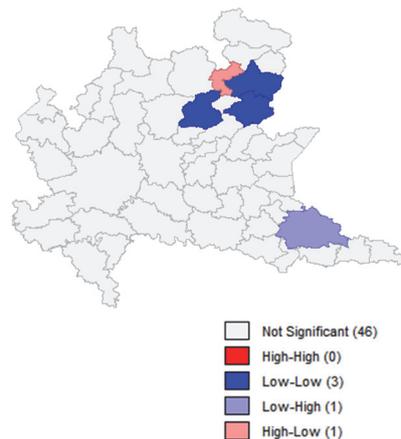


Figura 14.4 - “non profit verso la specializzazione”, 2011 (LISA cluster map)



Il terzo indice sintetico rappresenta “*il non profit della promozione e tutela dei diritti;*” che attraverso un sistema di servizi sociali e sanitari lavora per il raggiungimento del benessere della popolazione e la promozione dei diritti umani svolgendo attività di utilità sociale. Attraverso questo indice è possibile evincere un dettaglio maggiore sulle caratteristiche delle INP dedite al disagio, distinguendo fra quella componente del settore che svolge un ruolo importante nelle aree del welfare (assistenza sociale e sanità) più strutturata (per la presenza di lavoratori retribuiti) e orientata al market (Figura 14.5: Sistemi locali con valore dell'indicatore dell'ultimo quartile) da quella indirizzata a tipologie di disagio diverse dalla malattia (abusi e molestie, prostituzione, detenzione), che trova collocazione dove alta è la concentrazione territoriale di unità locali (Figura 14.5: Sistemi locali con valore dell'indicatore del primo quartile). L'analisi dell'autocorrelazione spaziale (Figura

¹⁵ INP che sostengono i propri costi attraverso entrate derivanti da contratti e/o convenzioni con le istituzioni pubbliche e ricavi derivanti da vendita di beni e servizi.

14.6) evidenzia quattro Sistemi locali che manifestano un alto livello di disagio raggiunto dai servizi delle istituzioni non profit. La *cluster map* illustra inoltre tre SI ugualmente caratterizzati dalla presenza di non profit dedito al disagio, ma confinanti con SI dove è presente un altro tipo di non profit.

Figura 14.5 - “non profit per il disagio”, 2011 (Box plot map)

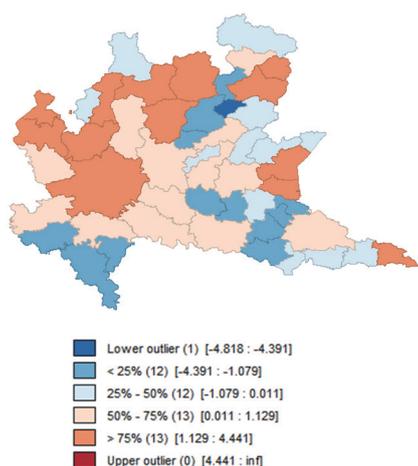
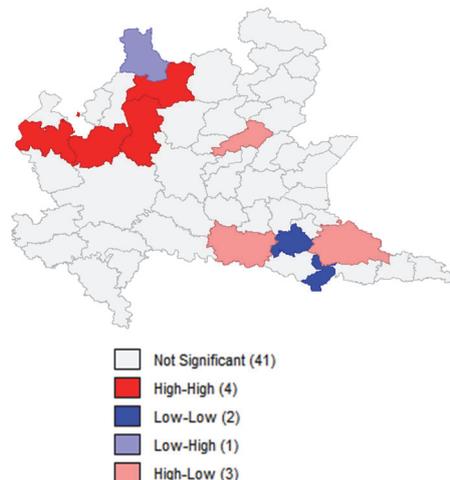


Figura 14.6 - “non profit per il disagio”, 2011 (LISA cluster map)



14.3.2 I SI della vulnerabilità familiare

L'analisi in componenti principali condotta sugli indicatori di vulnerabilità familiare ha portato alla selezione delle prime tre componenti (quelle con autovalore maggiore di 1) che rappresentano l'83,14 per cento della varianza.

La prima componente principale identifica il carico familiare ovvero il disagio delle famiglie dovuto principalmente alla loro condizione strutturale. Sono rappresentate due tipologie familiari descritte grazie agli indicatori strutturali: da una parte le famiglie con almeno un componente anziano di 65 anni e oltre che vivono con un reddito molto basso, dall'altra le famiglie con almeno un figlio minore, dai 0 ai 14 anni, che si affiancano alle famiglie numerose con 5 componenti e più.

Dalla *box-plot map* del carico familiare (Figura 14.7) emerge una configurazione abbastanza netta dei SI della Lombardia: i SI dell'ultimo quartile della distribuzione territoriale (quelli in cui l'indice di carico familiare risulta avere punteggi positivi) sono concentrati nei territori delle province di Brescia (SI di Montichiari, Chiari, Orzinuovi, Manerbio), Mantova (Castel Goffredo, Castiglione delle Stiviere), Bergamo (Grumello del Monte), Cremona, Sondrio nei quali è maggiore l'incidenza di famiglie con almeno un minore e molto numerose. Si tratta di luoghi metropolitani inseriti in una cornice industriale, che vede anche la forte presenza degli stranieri spesso associata agli indicatori di vulnerabilità appena presentati. I territori che ricadono nel primo quartile della distribuzione territoriale dell'indice, invece, sono prevalentemente quelli appartenenti ai SI di Vilminore di Scalve (Bergamo), Ponte di Legno (Brescia), Edolo (Brescia), Voghera (Pavia), Sondrio, Stradella (Pavia), Sermide (Mantova), Sondalo (Sondrio) e sono caratterizzati dalla forte incidenza di famiglie con almeno un componente anziano di 65 anni ed oltre e da redditi bassi. Si tratta di territori in prevalenza montani che vedono la forte presenza di residenti anziani.

Figura 14.7 - Carico familiare. 2012 (Box-plot map)

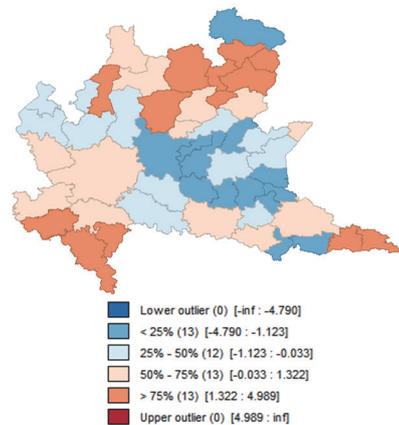
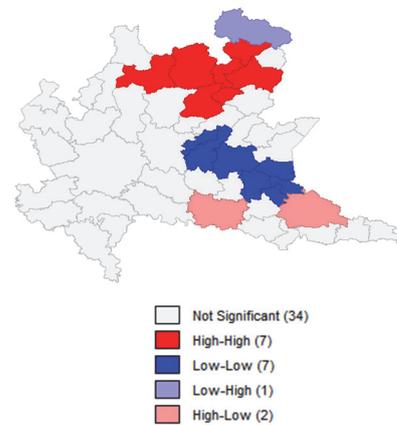


Figura 14.8 - Carico familiare. 2012 (LISA Cluster map)



La *cluster map* (Figura 14.8) individua un cluster composto da 7 sistemi locali di tipo *high-high*, che conferma la forte incidenza di vulnerabilità familiare nei territori montani dovuta alla presenza di popolazione anziana con bassi livelli di reddito. L'analisi dell'autocorrelazione spaziale evidenzia inoltre la diversità dei sistemi locali di Livigno, che si caratterizza per un valore alto dell'indicatore pur essendo vicino a territori che presentano un valore basso dello stesso, e di Cremona e Mantova che, al contrario, sono caratterizzati da un basso livello dell'indicatore di carico familiare.

La seconda componente principale descrive la povertà economica esplicitata da una parte dalla presenza di famiglie il cui capofamiglia non ha lavorato nemmeno un giorno nell'ultimo anno e non possiede redditi da pensione, dall'altra dalla presenza di redditi molto bassi¹⁶. Dalla *box-plot map* dell'indice (Figura 14.9) emerge la presenza di tre outlier identificati dai territori dei sistemi locali di Porlezza (Como), Luino (Varese) e Chiavenna (Sondrio) che presentano i valori estremi del polo negativo della componente, seguiti dai SI di Menaggio (Como), Sannazzaro de' Burgondi (Pavia), Chiari (Mantova), Edolo (Brescia), Orzinuovi (Mantova) e Tirano (Sondrio). I sistemi locali appartenenti all'ultimo quartile della distribuzione territoriale (quelli in cui l'indice di povertà economica familiare risulta avere punteggi positivi) sono quelli di Livigno, Lecco, Milano, Castiglione delle Stiviere, Mantova, Busto Arsizio, Bergamo, Pavia Brescia, Cremona, territori metropolitani ed industrializzati nei quali è forte la presenza delle famiglie straniere (con almeno un componente straniero o apolide) e delle famiglie numerose la cui condizione economica è vulnerabile.

L'analisi dell'autocorrelazione spaziale (Figura 14.10) mette in evidenza il sistema locale di Menaggio (valore alto dell'indicatore), nel quale l'incidenza delle famiglie con capofamiglia che non ha lavorato nell'ultimo anno è pari al 13,8 per cento (la media dei SI è 8,7 per cento).

¹⁶ Come già segnalato, i risultati relativi all'area insubrica sono condizionati dal fenomeno del lavoro oltre confine non disponibile nelle basi dati utilizzate (Cfr. Capitolo 3).

Figura 14.9 - Povertà economica familiare. 2012 (Box-plot map)

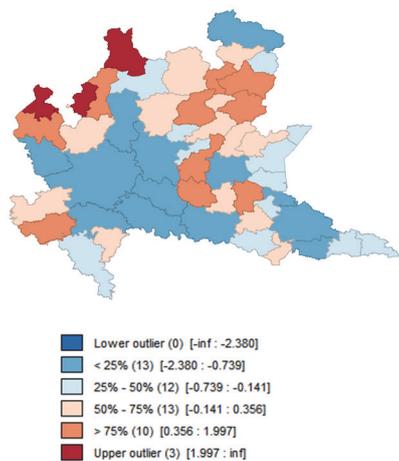
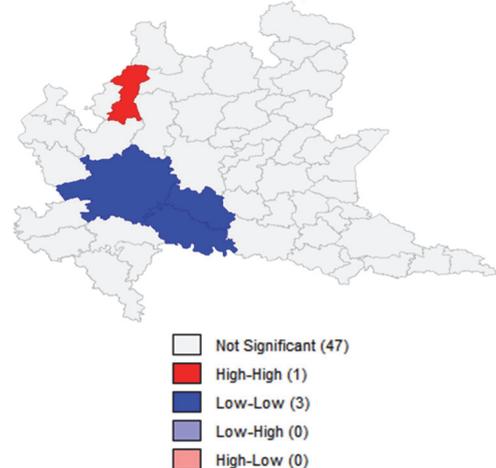


Figura 14.10 - Povertà economica familiare. 2012 (LISA cluster map)



La terza componente, infine, descrive la povertà lavorativa. Si ritrovano da una parte famiglie con almeno un componente tra i 18 e i 39 anni che non ha reddito né studia e dalle famiglie con capofamiglia che non ha lavorato almeno un giorno nell'anno e non ha reddito da pensione. In questo caso sono rilevanti le incidenze nei territori dei sistemi locali di Varese, Chiavenna, Como, Bergamo, Brescia, Luino, Desenzano del Garda, Livigno, Pavia e Milano per cui si può osservare che la mancanza di una occupazione e l'assenza di un posto nel mercato del lavoro è vissuta in misura maggiore dai capofamiglia dei centri urbanizzati e delle metropoli piuttosto che dalle famiglie residenti in altre zone. Allo stesso tempo, la terza componente è descritta anche dalle famiglie nelle quali il capofamiglia non ha alcun titolo di studio o ha la licenza elementare e dalle famiglie con almeno un anziano di 65 anni ed oltre. Queste tipologie insistono in prevalenza nella fascia est della regione, sui territori delle province di Mantova Cremona e Brescia (Figura 14.11), e, in particolare nei sistemi locali di Vestone (Brescia), Asola (Mantova), Sermide (Mantova), Poggio Rusco (Mantova), Casalmaggiore (Cremona), Suzzara (Mantova), Lumezzane (Brescia), Viadana (Mantova).

Figura 14.11 - Povertà lavorativa familiare. 2012 (Box-plot map)

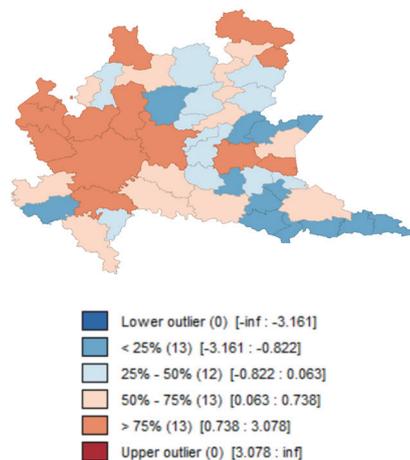
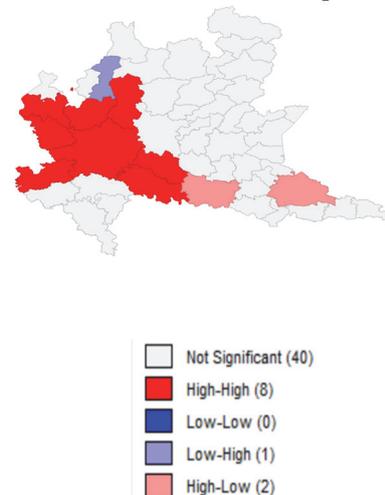


Figura 14.12 - Povertà lavorativa familiare. 2012 (LISA cluster map)



L'analisi dell'autocorrelazione spaziale (Figura 14.12) conferma la contrapposizione territoriale tra aree montane ed aree metropolitane come è possibile osservare dai SI evidenziati dal cartogramma: il cluster dei sistemi locali di tipo *high-high* di Milano, Crema e Lodi, è caratterizzato da territori che manifestano un alto livello dell'indice di povertà lavorativa. La *cluster map* permette inoltre di evidenziare anche la presenza di due SI ugualmente caratterizzati da un alto valore dell'indice ma in un caso – Cremona – il sistema locale è confinante con un sistema di tipo *high-high*, nell'altro – Mantova – è invece confinante con SI nei quali non è significativo il valore dell'indice di povertà lavorativa.

14.4 Uno sguardo di insieme

Una lettura combinata delle *Cluster map* (Figure 14.6 e 14.11) rivela il ruolo del non profit orientato a diverse tipologie di disagio (abusi e molestie, prostituzione, detenzione) in sistemi locali caratterizzati da povertà lavorativa familiare (dove alta è l'incidenza di famiglie con almeno un componente tra i 18 e i 39 anni che non ha reddito né studia e di famiglie con capofamiglia che non ha lavorato almeno un giorno nell'anno e non ha reddito da pensione). Negli stessi SI, con caratteristiche urbane, si evidenzia inoltre la presenza di istituzioni non profit di origini più recenti, attente ai bisogni e disagi emergenti (Figura 14.2). Viceversa nel versante montano, che risulta caratterizzato dalla forte incidenza di famiglie con almeno un componente anziano di 65 anni ed oltre e da redditi bassi, trova spazio il settore non profit specializzato in sanità e disagio psico-sociale (Figure 14.4 e 14.8). La lettura congiunta dei dati censuari ed amministrativi permette di analizzare, nei diversi contesti territoriali di riferimento, la relazione che esiste fra le attività svolte dalle istituzioni non profit e le diverse forme di disagio a cui esse si orientano. Il presente lavoro di carattere sperimentale, pone le basi per la costruzione e lo studio di una serie di indicatori che permettano di cogliere la risposta resiliente del settore ai bisogni sociali e che consentano di studiare la capacità delle istituzioni di dare supporto alla diverse tipologie di disagio assumendo forme differenziate in base alle esigenze espresse a livello territoriale.

CONCLUSIONI E PROSPETTIVE FUTURE

di Flavio Verrecchia

Il volume è il frutto di una rilevante attività di ricerca realizzata grazie alle conoscenze, esperienze e competenze maturate in diversi contesti territoriali e ambiti di ricerca e di produzione. È il prodotto della sinergia tra l'accademia, l'Istat e gli enti locali realizzatasi all'interno del progetto SPoT – dati e metodologie per lo sviluppo delle Statistiche per le Politiche Territoriali – avviato nel 2016 nell'ambito delle linee di attività della Direzione centrale per la raccolta dati dell'Istituto nazionale di statistica.

L'obiettivo principale della linea di attività ha riguardato la sperimentazione, con dati di fonte amministrativa, di approcci informativi di supporto alle politiche di livello locale e alle tipiche fasi di valutazione (*ex-ante*, *in itinere* ed *ex-post*). Insieme alle statistiche ufficiali, i dati di fonte amministrativa possono essere considerati – allo stato attuale in termini sperimentali – per la produzione di informazioni ad un livello di granularità sub-provinciale. Ad esempio, per il supporto delle politiche territoriali le statistiche sperimentali – pur potendo peccare di precisione nelle misure – possono certamente favorire la conoscenza dei contesti e delle tematiche oggetto di studio. Considerata questa prospettiva il progetto ha visto l'uso di strumenti di analisi e rappresentazione che per costruzione non fossero conflittuali con le statistiche ufficiali.

Proprio le attività realizzate, illustrate in questo volume, sono state sviluppate secondo la prospettiva delle statistiche sperimentali. Se da un lato questa prospettiva può non rispettare i criteri necessari all'introduzione stabile di nuove metodologie, alla sua traduzione in soluzioni tecnologiche e organizzative, al rispetto dei requisiti di qualità e delle regole di armonizzazione, dall'altro, ha un potenziale elevatissimo in quanto colma lacune conoscitive in maniera tempestiva e con periodicità frequente, dà impulso allo sviluppo di nuove analisi e di nuovi indicatori, garantisce un valido sostegno conoscitivo alle politiche.

Com'è stato detto nell'introduzione, la fonte dati sperimentale che è stata utilizzata in SPoT è *l'ARCHivio Integrato di Microdati Economici e Demografici* (ArchIMEDe), che prende il nome dall'omonimo progetto, che si è occupato di costruzione e aggiornamento di basi di dati per l'analisi territoriale nell'ambito del Sistema Integrato dei Microdati (Sim) dell'Istat, quali: Condizioni socioeconomiche delle famiglie; Popolazioni che usano un territorio e mobilità; Precarietà lavorativa. Queste basi di dati hanno consentito, come è avvenuto nella terza e quarta parte del testo, l'elaborazione di numerosi indicatori a livello comunale, sia relativi a importanti aspetti strutturali sia attinenti a fenomeni osservati anche nell'ambito della misurazione della Deprivazione e del Benessere Equo e Sostenibile nei Territori (*DeBEST*).

Il principale risultato atteso dalla realizzazione del volume non ha riguardato solo il rinnovamento dell'attenzione posta dall'Istat *nello sviluppo di un'approfondita conoscenza della realtà ambientale, economica e sociale dell'Italia ai diversi livelli territoriali e nel favorire i processi decisionali di tutti i soggetti della società* – che rappresenta una delle principali *mission* dell'Istituto – ma è anche consistito nella condivisione delle sperimentazioni raccolte che si

riferiscono a fasi diverse del processo di ricerca. Più in dettaglio, nel volume sono stati sviluppati contributi relativi a:

- dati, metadati e qualità delle fonti secondarie (Cfr. prima parte);
- approcci gestionali e informativi, tecniche per la costruzione di indicatori e metodi grafici rivolti alla diffusione di informazioni territoriali di supporto alle politiche dei governi locali (Cfr. seconda parte);
- applicazioni Sistan, dove risulta più forte il nesso tra pianificazione, programmazione e gestione, nell'ottica di soddisfare la necessità di statistiche a supporto di specifici processi decisionali pubblici (Cfr. terza parte);
- casi di studio su deprivazione e benessere equo e sostenibile nei territori ad un livello sub-provinciale (Cfr. quarta parte).

In prospettiva, i risultati delle linee di attività Istat attinenti alle statistiche sperimentali, di cui SPoT è un esempio, possono idealmente essere organizzati in quattro differenti tipologie già identificate dall'Istituto: classificazioni non standard ottenute a partire dalle tassonomie ufficiali utilizzate correntemente dall'Istat, oppure proposte in modo sperimentale nell'ambito di attività di analisi e ricerca attraverso elaborazioni condotte su microdati; nuovi indicatori prodotti sulla base dell'integrazione di una pluralità di fonti, ufficiali e non ufficiali, dove l'enfasi è posta sui fenomeni allo studio e non sulle fonti statistiche intese a descriverli; analisi e quadri interpretativi di fenomeni complessi ottenuti attraverso l'integrazione di fonti ufficiali; risultati di sperimentazioni su *big data*, caratterizzate per loro natura dall'utilizzo di fonti non ufficiali.

Grazie alla divulgazione dei risultati connessi a queste nuove sperimentazioni è possibile valorizzare il dato amministrativo estraendo informazioni fruibili da parte dei decisori locali e dell'utenza più in generale. In tal senso SPoT si è affiancato ad altre esperienze che miravano alla valorizzazione delle fonti di carattere sperimentale accanto ad altre più consolidate, come è avvenuto nel caso del sistema multi-fonte "A misura di comune", risultato della linea di attività "Misure di benessere e programmazione a livello comunale" avviata, anch'essa nel 2016, nell'ambito dei progetti della Direzione centrale per la pianificazione strategica, l'indirizzo del Sistema statistico nazionale, le relazioni istituzionali e gli affari internazionali (Dcps) dell'Istat. Rispetto a questo progetto, ad esempio, si è trattato di una sinergia più ampia. SPoT, infatti, ha condiviso *in itinere* i metodi di ottimizzazione implementati per lo sfruttamento a livello locale delle basi dati di ArchIMEDe (Cfr. capitolo 4) agevolando la presentazione di risultati intermedi (Bianchino e Ricci, 2017).

In conclusione le sperimentazioni descritte nel volume, sono in linea con il percorso intrapreso dall'Istat, da altri istituti di statistica stranieri e da Eurostat riguardo all'utilizzo di nuove fonti amministrative, all'applicazione di metodi innovativi nella produzione di dati e al più recente studio di adeguate modalità di diffusione delle statistiche sperimentali.

APPENDICE I – REDDITI

I prospetti che seguono riportano, per ognuna delle macro-voce di reddito le relative componenti¹ e la fonte da cui queste ultime sono derivate.

Prospetto A1.1 - Reddito da lavoro dipendente

| Fonte | Variabile | Operazione |
|-------------------------------|---|------------|
| BDR | redditi da lavoro dipendente e assimilato, al lordo degli straordinari. La popolazione di riferimento 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, ecc. | + |
| BDR | altri redditi assimilati a quelli di lavoro dipendente per i quali la detrazione non è rapportata al periodo di lavoro nell'anno, ad esclusione dell'assegno percepito dal coniuge. | + |
| BDR | premi di produzione assoggettati ad imposta sostitutiva (solo dipendenti del settore privato). | + |
| BDR | quota di reddito esente dei lavoratori frontalieri. I redditi prodotti in zone di frontiera sono infatti imponibili ai fini IRPEF per la parte eccedente ad una certa soglia. | + |
| TMNP | importi di alcuni trattamenti non pensionistici (p.es. disoccupazione, maternità) che, dal punto di vista fiscale, costituiscono redditi della stessa categoria di quelli sostituiti o perduti. | - |
| Archivio Lavoratori domestici | stima della retribuzione percepita dal lavoratore domestico, calcolata come il prodotto di retribuzione effettiva oraria e numero di ore retribuite per tutti i rapporti di lavoro domestico relativi allo stesso individuo. | + |
| Banca dati voucher | Somma degli importi riscossi nell'anno di riferimento per ogni prestazione. | + |

Prospetto A1.2 - Reddito da lavoro autonomo

| Fonte | Variabile | Operazione |
|-------|---|------------|
| BDR | reddito agrario. | + |
| BDR | reddito dell'impresa di allevamento. | + |
| BDR | reddito derivante dall'esercizio di arti e professioni. | + |
| BDR | redditi derivanti dalla utilizzazione economica, da parte dell'autore o inventore, di opere dell'ingegno, di brevetti industriali e di processi, ecc. | + |
| BDR | importo risultante dalla differenza tra l'ammontare lordo di una varietà di redditi e il totale delle spese, come i compensi derivanti da attività di lavoro autonomo, anche se svolte all'estero, non esercitate abitualmente, ecc. | + |
| BDR | reddito imponibile derivante da attività sportive dilettantistiche e prestazioni rese a favore di cori, bande musicali e filodrammatiche. | + |
| BDR | reddito d'impresa di spettanza dell'imprenditore al netto delle perdite d'impresa. | + |
| BDR | reddito d'impresa di spettanza dell'imprenditore al netto delle perdite d'impresa in contabilità ordinaria. | + |
| BDR | reddito di partecipazione in società esercenti attività d'impresa. | + |
| BDR | reddito di partecipazione in associazioni tra artisti e professionisti. | + |
| BDR | reddito di partecipazione in società semplici. | + |
| BDR | reddito da lavoro autonomo (da 770 semplificato) per cui non è presente la dichiarazione del percipiente. | + |
| UPF | reddito dei soggetti con regime dell'imprenditoria giovanile e lavoratori in mobilità, ovvero quella parte di contribuenti la cui attività d'impresa o professionale possiede determinati requisiti (sui ricavi, sul tipo di attività, ecc.). | + |

Prospetto A1.3 - Reddito da capitale reale

| Fonte | Variabile | Operazione |
|-------|--|------------|
| BDR | redditi da fabbricati soggetto a tassazione ordinaria (e sostitutiva). | + |
| BDR | reddito dominicale. | + |
| BDR | redditi relativi all'abitazione principale. | + |
| BDR | redditi relativi ai fabbricati non locati; reddito dominicale dei terreni non affittati. | + |

¹ Viene anche riportato, nella colonna "operazione" se la componente viene sommata o sottratta alla macro-voce di reddito a cui afferisce.

Prospetto A1.4 -Reddito da capitale finanziario

| Fonte | Variabile | Operazione |
|-------|---|------------|
| BDR | redditi di capitale. Sono inclusi gli utili e gli altri proventi equiparati in qualunque forma corrisposti da società di capitali o enti commerciali, residenti e non residenti e altri redditi di capitale, percepiti al lordo delle eventuali ritenute a titolo di acconto. | + |
| BDR | redditi da capitale imputati da Trust. | + |
| BDR | plusvalenze derivanti dalla cessione di partecipazioni qualificate. | + |
| BDR | imponibile delle plusvalenze derivanti dalla cessione di partecipazioni in società localizzati in territori con un regime fiscale privilegiato. | + |
| BDR | alcuni redditi a tassazione separata per i quali è stata effettuata l'opzione a tassazione ordinaria. | + |

Prospetto A1.5 - Reddito da pensione

| Fonte | Variabile | Operazione |
|-------|--|--|
| BDR | importi delle pensioni da lavoro e di reversibilità, dell'assegno di invalidità o delle pensioni di inabilità. Non sono inclusi nel reddito da pensione, in quanto esenti, le pensioni erogate agli invalidi civili, ai non vedenti e ai non udenti, le indennità di accompagnamento, le rendite (dirette) per infortunio sul lavoro e per le malattie professionali, le pensioni (dirette) di guerra, per assegno di benemerenzia ai perseguitati politici. | + |
| CP | importi delle pensioni indennitarie e di guerra. | + |
| CP | ammontare del reddito da pensioni non esenti (pensioni di vecchiaia, invalidità e superstiti). | + (se non presente importo pensionistico da BDR) |

Prospetto A1.6 - Reddito da trasferimenti

| Fonte | Variabile | Operazione |
|----------|--|------------|
| CP | reddito pensionistico derivante dalla/e pensione/i di invalidità civile e sociale/i. | + |
| TMNP | assegno al nucleo familiare pensionati delle gestioni dei lavoratori autonomi, assegno al nucleo familiare lavoratori agricoli, assegno al nucleo familiare lavoratori parasubordinati, assegno al nucleo familiare pensionati da lavoro dipendente, assegni al nucleo familiare dei comuni, assegni al nucleo familiare lavoratori domestici. | + |
| TMNP | assegni di maternità dei comuni e degli assegni di maternità dello stato. | + |
| Uniemens | assegni al nucleo familiare correnti per i lavoratori dipendenti del settore privato. | + |
| BDR | assegno periodico percepito dal coniuge o ex-coniuge. | + |

APPENDICE II – COMUNI FRONTALIERI

Prospetto A2.1 - Lista dei Comuni italiani di frontiera. Accordo tra la Svizzera e l'Italia relativo all'imposizione dei lavoratori frontalieri concluso il 3 ottobre 1974

| | | | |
|---------------------------------|-----------------------|--|-----------------------------|
| Sondrio | Luisago | Brissago-Valtravaglia | Vedano Olona |
| <i>Tutti i comuni</i> | Lurago d'Erba | Brunello | Veddasca ² |
| | Lurago Marinone | Brusimpiano | Venegono Inferiore |
| Como | Lurate Caccivio | Buguggiate | Venegono Superiore |
| Albavilla | Magreglio | Cadegliano-Viconago | Viggiù |
| Albese con Cassano | Mariano Comense | Cairate | Lecco |
| Albiolo | Masilianico | Canello | Abbadia Lariana |
| Alserio | Menaggio | Caravate | Bellano |
| Aizzate Brianza | Merone | Carnago | Bosisio Parini |
| Anzano del Parco | Mezzegra | Caronno Varesino | Casargo |
| Appiano Gentile | Moltrasio | Casale Litta | Cesana Brianza |
| Argegno | Monguzzo | Casalzuigno | Civate |
| Arosio | Montano Lucino | Casciago | Colico |
| Asso | Montemezzo | Cassano Magnago | Costa Masnaga |
| Barni | Montorfano | Cassano Valcuvia | Dervio |
| Bellagio | Mozzate | Castello Cabiaglio | Dorio |
| Bene Lario | Musso | Castelseprio | Esino Lario |
| Beregazzo con Figliaro | Nesso | Castelvecchiana | Introzio |
| Binago | Novedrate | Castiglione Olona | Lierna |
| Bizzarone | Olgiate Comasco | Castronno | Mandello del Lario |
| Blessagno | Oltrona di | Cavaria con Premezzo | Nibionno |
| Blevio | San Mamette | Cazzago Brabbia | Oliveto Lario |
| Bregnano | Orsenigo | Cislago | Pagnona |
| Brenna | Ossuccio | Cittiglio | Perledo |
| Brienno | Parè | Clivio | Rogeno |
| Brunate | Peglio | Cocquio-Trevisago | Sueglio |
| Bulgarograsso | Pellio Intelvi | Comabbio | Suello |
| Cabiato | Pianello del Lario | Comerio | Tremenico |
| Cadorago | Pigra | Cremenaga | Valmadrera |
| Caglio | Plesio | Crosio della Valle | Varenna |
| Cagno | Pognana Lario | Cusasso al Monte | Vendrogno |
| Campione d'Italia | Ponina | Cugliate-Fabiasco | Vestreno |
| Cantù | Ponte Lambro | Cunardo | |
| Canzo | Porlezza | Curigliana | Verbano Cusio Ossola |
| Capiago Intimiano | Proserpio | con Monteviasco | Antrona Schierano |
| Carate Urio | Pusiano | Cuveglia | Arizzano |
| Carbonate | Ramponio Verna | Cuvio | Aurano |
| Carimate | Rezzago | Daverio | Baceno |
| Carlazzo | Rodero | Dumenza | Bannio Anzino |
| Carugo | Ronago | Duno | Bee |
| Casasco d'Intelvi | Rovellasca | Fagnano Olona | Beura-Cardezza |
| Casolino d'Erba | Rovello Porro | Ferrera di Varese | Bognanco |
| Casnate con Bernate | San Bartolomeo | Gallarate | Calasca-Castiglione |
| Cassina Rizzardi | Val Cavargna | Galliate Lombardo | Cambiasca |
| Castelmarte | San Fedele Intelvi | Gavirate | Cannero Riviera |
| Castelnuovo Bozzente | San Fermo della | Gazzada Schiavio | Cannobio |
| Castiglione d'Intelvi | Battaglia | Gemonio | Caprezzo |
| Cavallasca | San Nazzaro | Gemignaga | Cavaglio-Spocchia |
| Cavargna | Val Cavargna | Goria Maggiore | Ceppo Morelli |
| Cerano d'Intelvi | Val Cavargna | Goria Minore | Cossogno |
| Cermenate | Sala Comacina | Gornate-Olona | Crevoladossola |
| Cernobbio | San Siro ² | Grantola | Creodo |
| Cirimido | Schignano | Inarzo | Cursolo-Orasso |
| Civerna | Semina Comasco | Inaduno Olona | Domodossola |
| Claino con Osteno | Solbiate | Jerago con Orago | Druogno |
| Colonno | Sorico | Lavena Ponte Tresa | Falmenta |
| Como | Sormano | Laveno-Mombello | Formazza |
| Corrido | Stazzona | Leggiano | Ghiffa |
| Crema | Tavernerio | Lonate Ceppino | Curro |
| Cucciago | Torno | Lozza | Intragna |
| Cusino | Tremezzo | Luino | Macugnaga |
| Dizzasco | Trezzone | Luvinate | Malesco |
| Domaso | Turate | Maccagno ³ | Masera |
| Dongo | Uggiate-Trevano | Malgesso | Mergozzo |
| Dosso del Liro | Val Rezzo | Malnate | Miazzina |
| Drezzo | Valbrona | Marchirolo | Montecrestese |
| Erba | Valmorea | Marzio | Montescheno |
| Eupilio | Valsolda | Masciago Primo | Oggebbio |
| Faggeto Lario | Veleso | Mercallo | Pallanzeno |
| Faloppio | Veniano | Mesenzana | Piedimulera |
| Fenegrò | Vercana | Montegrino Valtravaglia | Pieve Vergonte |
| Figino Serenza | Vertemate | Monvalle | Premeno |
| Fino Mornasco | con Minoprio | Morazzone | Premia |
| Garzeno | Villa Guardia | Mornago | Premosello-Chiovenda |
| Gera Lario | Zelbio | Oggiona con | Re |
| Gironico | Varese | Santo Stefano | San Bernardino |
| Grandate | Agra | Orino | Verbano |
| Grandola ed Uniti | Albizzate | Osmate ⁴ | Santa Maria Maggiore |
| Gravedona ed Uniti ¹ | Arcisate | Pino sulla Sponda del Lago Maggiore ³ | Seppiana |
| Griante | Arsago Seprio | Porto Ceresio | Toceno |
| Guanzate | Azzate | Porto Valtravaglia | Trarego Viggiona |
| Inverigo | Azzio | Rancio Valcuvia | Trasquera |
| Laglio | Barasso | Saltrio | Trontano |
| Laino | Bardello | Sangiano | Valstrona |
| Lamburgo | Bedero Valcuvia | Solbiate Arno | Vanzone con San Carlo |
| Lanzo d'Intelvi | Besano | Solbiate Olona | Varzo |
| Lasnigo | Besnate | Sumirago | Verbania |
| Lenno | Besozzo | Ternate | Viganella |
| Lezzeno | Biandronno | Tradate | Vignone |
| Limido Comasco | Bisuschio | Travedona-Monate | Villadossola |
| Lipomo | Bodio Lomnago | Tronzano Lago | Villette |
| Livo | Bregano | Maggiore | Vogogna |
| Locate Varesino | Brenta | Valganna | |
| Lomazzo | Brezzo di Bedero | Varano Borghi | |
| Longone al Segrino | Brinzio | Varese | |

¹ **Gravedona ed Uniti**: costituito con Legge regionale n. 1 del 10.02.2011, dalla fusione di Consiglio di Rumo, Gemmasino e Gravedona. ² **San Siro**: ricostituito con Legge regionale n. 29 del 29.11.2002, dalla fusione di Sant'Abbondio e Santa Maria Rezzonico, era stato soppresso nel 1928. ³ **Maccagno con Pino e Veddasca**: costituito dal 4.02.2014, dalla fusione di Maccagno, Pino sulla sponda del Lago Maggiore e Veddasca. ⁴ **Osmate**: si chiamava Osmate Lentate prima del 1970.

Convenzioni¹

Convenzione Italia – Svizzera. Podestà impositiva esclusiva del Paese in cui i redditi sono prodotti. La zona di confine riguarda i comuni italiani delle province di Como, Varese, Lecco e del Verbano-Cusio-Ossola situati nella fascia entro 20 km dal confine elvetico ed i comuni della provincia di Sondrio, mentre per la Svizzera sono considerati frontalieri i territori del Cantone Ticino, il Cantone Vallese ed il Cantone dei Grigioni. Con riferimento, quindi, ai rapporti con la Svizzera, in riferimento alla tassazione sui redditi, la nozione di frontaliere riguarda solo i lavoratori dipendenti che quotidianamente si recano dalla propria residenza, sita in un comune prossimo al confine, in uno dei Cantoni svizzeri confinanti con l'Italia. In tal caso, il lavoratore italiano che si reca in Svizzera è fiscalmente esente in Italia e tassato nella Confederazione elvetica. Per i frontalieri italiani oltre la fascia di 20 chilometri (n.b. di riferimento i comuni inseriti negli accordi), invece, è applicabile la disciplina di una franchigia e del credito d'imposta.

Convenzione Italia – Francia. Potestà impositiva esclusiva del Paese in cui il soggetto è residente. È applicabile la disciplina della franchigia. Le regioni interessate in Italia sono: Valle d'Aosta, Piemonte, Liguria. In Francia i dipartimenti coinvolti sono: Alta Savoia, Savoia, Alte Alpi, Alpi dell'Alta Provenza, Alpi Marittime.

Comuni frontalieri con la Svizzera e questionario ridotto del censimento

I comuni che hanno svolto il Censimento Popolazione 2011 principalmente con i fogli di famiglia in forma breve (oltre che con il foglio classico) erano quelli per i quali al 1° gennaio 2008 era stata calcolata una popolazione residente di almeno 20mila abitanti. Documenti di interesse sono:

- la Circolare n. 1 della Rilevazione dei Numeri Civici, che contiene l'elenco complessivo dei suddetti comuni;
- il Manuale della rilevazione del 15° Censimento Generale della Popolazione e delle Abitazioni – pag. 4 – dove si trova esplicito riferimento al fatto che il questionario in forma ridotta sarebbe stato impiegato, in parte, solo nei comuni con almeno 20mila abitanti al 1° gennaio 2008.

Prospetto A2.2 - Residenti nei comuni con almeno 20mila abitanti delle province frontaliere con la Svizzera (Varese, Como, Sondrio, Verbano-Cusio-Ossola) al 1° gennaio 2008

| Classe ampiezza demografica | Provincia | Codice Comune | Comuni | Totale residenti |
|-----------------------------|----------------------------|---------------|-----------------|------------------|
| >=50.000 | 013 - Como | 013075 | Como | 83.175 |
| | 012 - Varese | 012133 | Varese | 82.037 |
| | 012 - Varese | 012026 | Busto Arsizio | 80.633 |
| | 012 - Varese | 012070 | Gallarate | 50.156 |
| 20.000-49.999 | 012 - Varese | 012119 | Saronno | 38.126 |
| | 013 - Como | 013041 | Cantù | 37.824 |
| | 103 - Verbano-Cusio-Ossola | 103072 | Verbania | 30.943 |
| | 013 - Como | 013143 | Mariano Comense | 23.043 |
| | 014 - Sondrio | 014061 | Sondrio | 22.214 |
| | 012 - Varese | 012040 | Cassano Magnago | 21.239 |

¹ Per aggiornamenti normativi fare riferimento agli accordi bilaterali (online: <https://www.eda.admin.ch/dea/en/home/bilaterale-abkommen/ueberblick.html>).

APPENDICE III – MISSING DATA

Per aggregazioni di dati territoriali finalizzate al calcolo di indicatori, è di interesse verificare il peso dei *missing*. Di utilità, anche la verifica dell'omogeneità di comportamento rispetto al territorio di riferimento, impiegando ad esempio indicatori costruiti sull'ultimo anno disponibile o con fonti alternative (p.es. censimento). Questo tipo di *missing* dipende principalmente dal ritardo nella fornitura delle LAC.

Prospetto A3.1 - Missing. Lombardia 2012 e 2013 (valori assoluti)

| Codice prov. | Denominazione provincia | PROCOM | Denominazione comune | Comuni 2012 | Comuni 2013 | Pop. 2012 (dati 2011) | Pop. 2013 (dati 2011) |
|--------------|-------------------------|--------|------------------------------|-------------|-------------|-----------------------|-----------------------|
| 012 | Varese | 012019 | Brenta | | 1 | | 1.798 |
| 012 | Varese | 012035 | Caronno Varesino | | 1 | | 4.940 |
| 012 | Varese | 012142 | Maccagno con Pino e Veddasca | 1 | 1 | 2.444 | 2.444 |
| 013 | Como | 013037 | Caglio | 1 | 1 | 430 | 430 |
| 013 | Como | 013099 | Faloppio | | 1 | | 4.274 |
| 013 | Como | 013159 | Mozzate | | 1 | | 8.208 |
| 013 | Como | 013179 | Pellio Intelvi | 1 | | 1.004 | |
| 013 | Como | 013183 | Pianello del Lario | 1 | | 1.041 | |
| 013 | Como | 013193 | Pusiano | | 1 | | 1.327 |
| 013 | Como | 013212 | Senna Comasco | 1 | | 3.171 | |
| 013 | Como | 013217 | Sormano | 1 | | 633 | |
| 013 | Como | 013250 | Bellagio | 1 | 1 | 3.796 | 3.796 |
| 013 | Como | 013251 | Colverde | 1 | 1 | 5.234 | 5.234 |
| 013 | Como | 013252 | Tremezzina | 1 | 1 | 5.078 | 5.078 |
| 014 | Sondrio | 014061 | Sondrio | 1 | | 21.642 | |
| 015 | Milano | 015035 | Bubbiano | 1 | | 2.215 | |
| 015 | Milano | 015115 | Lacchiarella | | 1 | | 8.390 |
| 015 | Milano | 015246 | Zelo Surrigone | 1 | | 1.477 | |
| 015 | Milano | 015250 | Baranzate | | 1 | | 10.779 |
| 016 | Bergamo | 016019 | Barbata | 1 | 1 | 698 | 698 |
| 016 | Bergamo | 016027 | Blello | | 1 | | 76 |
| 016 | Bergamo | 016031 | Bonate Sotto | | 1 | | 6.455 |
| 016 | Bergamo | 016033 | Bossico | | 1 | | 969 |
| 016 | Bergamo | 016048 | Camerata Cornello | | 1 | | 627 |
| 016 | Bergamo | 016061 | Cassiglio | | 1 | | 122 |
| 016 | Bergamo | 016062 | Castelli Calepio | | 1 | | 9.612 |
| 016 | Bergamo | 016098 | Filago | 1 | | 3.176 | |
| 016 | Bergamo | 016110 | Gaverina Terme | 1 | | 871 | |
| 016 | Bergamo | 016172 | Pontirolo Nuovo | 1 | | 4.993 | |
| 016 | Bergamo | 016185 | Roncola | | 1 | | 739 |
| 016 | Bergamo | 016252 | Sant'Omobono Terme | 1 | 1 | 3.893 | 3.893 |
| 016 | Bergamo | 016253 | Val Brembilla | 1 | 1 | 4.522 | 4.522 |
| 017 | Brescia | 017105 | Marmentino | | 1 | | 676 |
| 017 | Brescia | 017128 | Ossimo | | 1 | | 1.442 |
| 017 | Brescia | 017139 | Pertica Alta | | 1 | | 599 |
| 017 | Brescia | 017140 | Pertica Bassa | | 1 | | 686 |
| 017 | Brescia | 017153 | Preseglie | | 1 | | 1.577 |
| 017 | Brescia | 017176 | Sellero | | 1 | | 1.506 |
| 017 | Brescia | 017194 | Valvestino | | 1 | | 212 |
| 018 | Pavia | 018039 | Castello d'Agogna | 1 | | 1.091 | |
| 018 | Pavia | 018075 | Gravellona Lomellina | 1 | | 2.688 | |
| 018 | Pavia | 018154 | Suardi | 1 | | 651 | |
| 018 | Pavia | 018168 | Valle Lomellina | 1 | | 2.153 | |
| 018 | Pavia | 018172 | Velezzo Lomellina | 1 | | 101 | |
| 018 | Pavia | 018191 | Cornale e Bastida | 1 | 1 | 895 | 895 |
| 018 | Pavia | 018192 | Corteolona e Genzone | 1 | 1 | 2.490 | 2.490 |
| 019 | Cremona | 019027 | Castelvisconti | 1 | 1 | 330 | 330 |
| 019 | Cremona | 019033 | Corte de' Frati | | 1 | | 1.423 |
| 019 | Cremona | 019050 | Grontardo | | 1 | | 1.469 |
| 019 | Cremona | 019063 | Olmeneta | | 1 | | 967 |
| 019 | Cremona | 019100 | Spinadesco | | 1 | | 1.575 |
| 019 | Cremona | 019110 | Trigolo | | 1 | | 1.760 |

Segue ->

Prospetto A3.1 (segue) - Missing. Lombardia 2012 e 2013 (valori assoluti)

| Codice prov. | Denominazione provincia | PROCOM | Denominazione comune | Comuni 2012 | Comuni 2013 | Pop. 2012 (dati 2011) | Pop. 2013 (dati 2011) |
|--------------------|-------------------------|--------|----------------------|-------------|-------------|-----------------------|-----------------------|
| 020 | Mantova | 020007 | Bozzolo | | 1 | | 4.181 |
| 020 | Mantova | 020071 | Borgo Virgilio | 1 | 1 | 14.400 | 14.400 |
| 097 | Lecco | 097012 | Calco | 1 | | 5.113 | |
| 097 | Lecco | 097078 | Suello | 1 | | 1.686 | |
| 097 | Lecco | 097091 | Verderio | 1 | 1 | 5.659 | 5.659 |
| 097 | Lecco | 097092 | La Valletta Brianza | 1 | 1 | 4.676 | 4.676 |
| Lombardia | | | | 31 | 41 | 108.251 | 130.934 |
| Lombardia % | | | | 2% | 3% | 1% | 1% |

Fonte: Elaborazioni su dati ArchIMEDe e Censimento della popolazione 2011.

RIFERIMENTI BIBLIOGRAFICI

Prefazione

Demchenko Y., Membrey P., deLaat C. (2014) Defining architecture components of the Big Data Ecosystem. In Proceedings of the 2014 International Conference on Collaboration Technologies and Systems (CTS 2014), May 19-23, 2014, Minneapolis, MN, USA.

Introduzione al volume

Bianchino A. e Garofalo G. (2017) Un sistema integrato di indicatori per le politiche locali” Convegno Uschi, Trento, Settembre 2017. Disponibile online <http://www.uschi.it/web/contenuti/speciali/1357-speciale-convegno-uschi-trento-14-e-15-settembre-2017>.

Capitolo 1

Ambroselli S. (2015) I codici identificativi univoci all’interno del SIM (Sistema Integrato di Microdati). Istat, Working Papers, 5.

Boarini R., Oliveira Martins J., Strauss H., de la Maisonneuve C. and Nicoletti G. (2008) Investment in tertiary education: main determinants and implications for policy. CESifo Economic Studies, 54(2), 277-312.

Cancian M., and Reed D. (2002) Changes in family structure: implications for poverty and related policy. In Danziger S.H., Haveman R.H. (eds.) Understanding poverty (pp. 69-96). Russell Sage Foundation Books at Harvard University Press.

De Looper M. and Lafortune G. (2009) Measuring disparities in health status and in access and use of health care in OECD Countries. OECD Health Working Paper, No. 43, Paris.

Donati E. (1991) Vivere sole in aree metropolitane. In Donati E. (eds.) Franco Angeli. Milano.

Garofalo G. (2014) Il Progetto ArchIMEDe obiettivi e risultati sperimentali. Istat Working Papers, 9.

Marino M.R. and Zizza R. (2012) The personal income tax evasion in Italy: an estimate by taxpayer’s type. In: Pickhardt M., Prinz A. (eds.), Tax evasion and the shadow economy, Edward Elgar Publishing. 33-60.

McLanahan S. (1985) Family structure and the reproduction of poverty. American Journal of Sociology, 90(4), 873-901.

Miyamoto K. and Chevalier A. (2010) Education and health, Chapter 4 of Improving Health and Social Cohesion through Education. OECD Publishing.

OECD (2010) The OECD innovation strategy. Getting a head start on tomorrow, OECD Publishing, Paris.

Sianesi B. and Van Reenen J. (2003) The returns to education: macroeconomics. Journal of Economic Surveys, 17(2), 157-200.

Wallgren A. and Wallgren B. (2014) Register-based statistics: statistical methods for administrative data, John Wiley & Sons, Ltd.

Unece (2011) Canberra Group Handbook on Household Income Statistics. Geneva.

Capitolo 2

- Dion M. (2007) *Metadata: an Integral Part of Statistics Canada's Data Quality Framework*, Statistics Canada. Canada.
- Dippo C.S., Sundgren B. (2012) *The Role of Metadata in Statistics*, Bureau of Labor Statistics.
- Eurostat (2015) *ESS Quality Assurance Framework, version 1.2*, Eurostat.
- Fröschl K. (1997) *Metadata Management in Statistical Information Processing*, Springer-Verlag Wien.
- Iwig W., Berning M., Marck P., Prell M. (2013) *Quality Assessment Tool for Administrative Data*. Subcommittee's Data Quality Working Group - Federal Committee on Statistical Methodology (FCSM). Data.
- Wallgren A. and B. Wallgren (2011) *Quality Assessment of Administrative Data. Research and Development – Methodology reports from Statistics Sweden 2011:2*.
- Wallgren A. and Wallgren B. (2014) *Register-based statistics: statistical methods for administrative data*, John Wiley & Sons, Ltd.
- United Nations Economic Commission for Europe Conference of European statisticians, European Commission Statistical Office of the European Union (Eurostat), Organisation for Economic Co-operation and Development (OECD) Statistics Directorate (2010) *Work Session on Statistical Metadata (METIS) - Final Report of the meeting*. Springer-Verlag. Wien.

Capitolo 3

- Ballabio S., Comune M.E., Ferrazza D., Verrecchia F., Vitalini A. (2016a) *La statistica ufficiale e le decisioni per il territorio. Dalle indagini alla sperimentazione su dati amministrativi. Caso di studio: Archimede e gli Early School Leavers*. Convegno di approfondimento: Il sistema educativo Lombardo: sfide e scenari di sviluppo. Le azioni di conoscenza e contrasto della dispersione scolastica e formativa in Lombardia. Éupolis Lombardia, Presso la Regione Lombardia, 12 maggio 2016.
- Ballabio S., Comune M.E., Ferrazza D., Verrecchia F., Vitalini A., Viviano L. (2016b) *Le lenti del territorio per la qualità delle statistiche da fonte amministrativa*. Convegno MILEs2016 (Milano - Impresa, Lavoro e Società 2016). Presso l'Università degli Studi di Milano Bicocca, 5 ottobre 2016. Poster.
- Ballabio S., Comune M.E., Ferrazza D., Verrecchia F., Vitalini A., Viviano L. (2016c) *Chi sono i NEET? Un'analisi a partire dai dati ArchIMEDe Neeting*. Convegno nazionale sui NEET. Istituto G. Toniolo di Studi Superiori, Fondazione Cariplo, Università Cattolica del Sacro Cuore Milano, Milano, 3-4 novembre 2016. Poster.
- Comune M.E., Viviano L. (2016) *La stima del lavoro frontaliero in Lombardia*, Dodicesima conferenza italiana di Statistica. Poster e slide.
- Comune M.E., Ronconi S., Viviano L. (2017) *The challenge of permeable borders: estimating cross-border employment between Italy and Switzerland*. In "Selected papers from the 2016 Conference of European statistics stakeholders". Eurostat. pp. 50-59. ISBN: 978-92-79-73629-2.
- Ferrazza D., Ballabio S., Comune M.E., Verrecchia F., Vitalini A., Viviano L. (2017) *Chi sono i NEET? Un'analisi a partire dai dati ARCHIMEDE*. Vita e Pensiero, Quaderni, 6 – Rapporto Giovani: Una generazione in panchina da Neet a

- risorsa per il Paese. A cura di Sara Alfieri, Emiliano Sironi. pp. 58-68. Ebook: ISBN 978-88-343-3324-2.
- Ballabio S., Carra A., Casacci S., Ferrazza D., Verrecchia F., Vitalini A., Viviano L. (2018a) In: Journées de Méthodologie Statistique (JMS2018) - INSEE, Paris, 12-14 giugno 2018.
- Ballabio S., Carra A., Casacci S., Ferrazza D., Verrecchia F., Vitalini A., Viviano L. (2018b) An experiment with Early School Leavers to measure the quality of integrated administrative data used to support local policies. In: European Conference on Quality in Official Statistics (Q2018) – Eurostat, Statistics Poland, Krakow, 26-29 giugno 2018.

Introduzione alla parte II: Metodi

- Inmon W. H. (1996) Building the Data Warehouse (2nd Edition), Wiley, New York.
- Anselin L. (1994) Exploratory spatial data analysis and geographic information systems, in M. Painho (eds.), New tools for spatial analysis, EUROSTAT, Luxembourg, pp. 45-54.
- Anselin L. (2004) Exploring spatial data with GeoDaTM: a workbook, University of Illinois, Urbana.

Capitolo 4

- Chen Peter P.S. (1976) The entity-relationship model: a basis for the enterprise view of data, ACM New York, NY, USA.
- Inmon W. H. (1996) Building the Data Warehouse (2nd Edition), Wiley, New York.
- Istat (1989) Manuale di tecniche d'indagine. 2 - il questionario: progettazione, redazione e verifica. Note e relazioni n. 1. Roma. ISSN: 0535-9856.

Capitolo 5

- Aureli Cutillo E (1996) Lezioni di statistica sociale. Parte seconda, sintesi e graduatorie. CISU, Roma.
- Delvecchio F. (1995) Scale di misura e indicatori sociali. Cacucci Editore, Bari.
- Dunteman G. H (1989) Principal Components Analysis. Sage Publications, Newbury Park.
- Harbison F. H., Maruhnic, J., Resnick, J. R. (1970) Quantitative Analyses of Modernization and Development. Princeton University Press, New Jersey.
- De Muro P., Mazziotta M., Pareto A. (2011) Composite Indices of Development and Poverty: An Application to MDGs. Soc. Indic. Res., 104: 1-18.
- Mazziotta M., Pareto A. (2011) Un indice sintetico non compensativo per la misura della dotazione infrastrutturale: un'applicazione in ambito sanitario. Rivista di Statistica Ufficiale, 1/2011, 63-79.
- Mazziotta M., Pareto A. (2016) "On a Generalized Non-compensatory Composite Index for Measuring Socio-economic Phenomena". Social Indicators Research, Vol. 127, Issue 3, pp 983-1003, Springer.
- Mazziotta C., Mazziotta M., Pareto A., Vidoli F. (2010) La sintesi di indicatori territoriali di dotazione infrastrutturale: metodi di costruzione e procedure di ponderazione a confronto. Review of Economics and Statistics for Regional Studies, 1.

- Munda G., Nardo M. (2005) Non-Compensatory Composite Indicators for Ranking Countries: A Defensible Setting. European Commission-JRC, Ispra.
- Palazzi P. (2004) Lo sviluppo come fenomeno multidimensionale. Confronto tra l'ISU e un indice di sviluppo sostenibile. *Moneta e Credito*, 227.
- OECD (2008) Handbook on Constructing Composite Indicators. Methodology and user guide. OECD Publications, Paris.

Capitolo 6

- Anselin L. (1994) Exploratory spatial data analysis and geographic information systems, in M. Painho (eds.), *New tools for spatial analysis*, EUROSTAT, Luxembourg, pp. 45-54.
- Anselin, L. (1995) Local Indicators of Spatial Association—LISA. *Geographical Analysis*, 27, pp. 93–115.
- Anselin L. (2004) *Exploring spatial data with GeoDaTM: a workbook*, University of Illinois, Urbana.
- Anselin, L., Syabri, I. and Kho, Y. (2006) GeoDa: An Introduction to Spatial Data Analysis. *Geographical Analysis*, 38, pp. 5–22.
- Devine, O. J., Louis, T. A. and Halloran, M. E. (1994) Empirical bayes estimators for spatially correlated incidence rates. *Environmetrics*, 5, pp. 381–398.

Capitolo 7

- Riva L., Trentini M. (1998) *Informazione statistica e processi decisionali pubblici: un quadro di riferimento*, Quaderno n. 145, Rapporti di ricerca del Dipartimento Metodi quantitativi, Università degli Studi di Brescia, Brescia.

Capitolo 8

- Garofalo G. (2014) *Il progetto ArchIMEDe: obiettivi e risultati sperimentali*, ISTAT, Working papers 9.
- Filippini R., Laghi A. e Ricciari V. (2016) *Percorsi di istruzione, formazione, abbandono e inserimento lavorativo. Sperimentazione di integrazione delle fonti regionali, comunicazione alla 12a Conferenza nazionale di statistica*, Roma.
- Filippini R., Laghi A. e Ricciari V. (2017) *Giovani senza segnali di istruzione, formazione, lavoro in Emilia Romagna e Lombardia in Rapporto Giovani. Una generazione in panchina. da Neet a risorsa per il paese*, a cura di Alfieri S. e Sironi E., Vita e Pensiero, Istituto Toniolo.
- Fossi L. e Ricciari V. (2015) "Evoluzione dei SLL in lombardia e i professionisti nei SLL fra 2007 e 2012", in *MILeS2015 | Milano - Impresa, Lavoro e Società 2015*, ISTAT, 2 ottobre 2015.
- Rappelli F. e Ricciari V. (2017) *L'uso dei dati amministrativi a supporto della programmazione delle politiche in Lombardia: l'analisi delle comunicazioni obbligatorie come strumento per adeguare l'offerta formativa IeFP alla domanda di lavoro*. XXXVIII Conferenza italiana di scienze regionali (AISRE).

Introduzione alla parte IV: Applicazioni sperimentali su dati amministrativi

- Autigna L. P. e Filandri M. (2015) *L'approccio multidimensionale alla povertà: le residenze temporanee della Regione Piemonte*, in «Territorio», 75, pp. 70-6.
- Lucchini M., Pisati M. e Schizzerotto A. (2007) *Stati di deprivazione e di benessere nell'Italia contemporanea. Un'analisi multidimensionale*, in Brandolini, C. e

- Saraceno, C. (a cura di), *Povert  e Benessere. Una geografia delle disuguaglianze in Italia*, Bologna, Il Mulino, pp. 271-303.
- Morlicchio E. (2012) *Sociologa della povert *, Bologna, Il Mulino.
- Struffolino E. e Filandri M. (2015) *Accumulare povert ? Note sul Report dell'ISTAT "La povert  in Italia 2014"*, in «Menab  di Etica e Economia», 27.
- Townsend, P. (1987) *Poverty and Labour in London. Interim Report*, London, The Low Pay Unit.

Capitolo 9

- Carrieri V. (2012) I working poor in Italia: quanti sono, chi sono, quanto sono poveri. *Rivista delle Politiche Sociali/Italian Journal of Social Policy*, 2/2012.
- Filandri M., Struffolino E. (2012) Lavoratori con basso reddito o occupati in famiglie povere? Un'analisi del fenomeno in Italia prima e dopo la crisi. Convegno Nazionale AIS-ELO, Cause e impatto della crisi. Individui, territori, istituzioni. Universit  della Calabria.
- Lucifera C. (2014) I working poor. Un'analisi dei lavoratori a basso reddito dopo la crisi. Cnel, 1 luglio 2014.
- Franzini M. (2017) Gli effetti della crisi sulla disuguaglianza economica. Convegno Istat, Universit  di Roma Sapienza. 26 novembre 2017.
- Crettaz E., Farine A. (2009) Salari bassi e *working poor*, in USTAT, Dati Statistiche e societ , A IX, n.1.
- Fulvimari A., Piroli G., Tanay F., Vandeppla A. Working (2017) lives the foundation of prosperity for all generation.
- Istat (2017) La povert  in Italia, in Statistiche report, 26 giugno 2018, Istat, Roma.

Capitolo 10

- Cadum E., Costa G., Biggeri A., Martuzzi M. (1999), Deprivazione e mortalit : un indice di deprivazione per l'analisi delle disuguaglianze su base geografica. *Epidem Prev*, 23: 175-187.
- Caranci N., Biggeri A., Grisotto L., Pacelli B., Spadea T., Costa G. (2010) L'indice di deprivazione italiano a livello di sezione di censimento: definizione, descrizione e associazione con la mortalit . *Epidemiol Prev*; 34: 167-176.
- Caranci N., Costa G. (2009) Un indice di deprivazione a livello aggregato da utilizzare su scala nazionale: giustificazione e composizione dell'indice in Costa G, Cislaghi C., Caranci N. (a cura di) "Disuguaglianze sociali di salute. Problemi di definizione e misura", Salute e societ , Franco Angeli.
- Costa G., Cislaghi C., Zocchetti C. et al (2011) Esistono misure valide dell'effetto della deprivazione socioeconomica sul fabbisogno sanitario? Alcune indicazioni metodologiche per il Servizio sanitario nazionale italiano. In *Politiche sanitarie*, 2011, 12, 3.
- Garofalo G. (2014) Il Progetto ArchIMEDe. Obiettivi e risultati sperimentali. Istat Working Papers n. 9, 2014.
- Massoli P., Mazziotta M., Pareto A., Rinaldelli C. (2015) COMIC: a tool for composite indices evaluation. Conference Dealing with complexity in society: from plurality of data to synthetic indicators, Padova, 17-18 Settembre.
- Mazziotta M., Pareto A. (2016) On a Generalized Non-compensatory Composite Index for Measuring Socio-economic Phenomena. In *Social Indicators Research*, 127: 983-1003.

- Pasetto R., Caranci N., Pirastu N. (2011) L'indice di deprivazione negli studi di piccola area su ambiente e salute, *Epidemiol e Prev* 35 (5-6) Suppl. 4: 1-204.
- Rosano A., Caranci N., De Felici P. et al. (2016), Utilizzo degli indici di deprivazione per orientare le politiche pubbliche di contrasto alla povertà, *Isfol OA*: <http://isfoloa.isfol.it/xmlui/handle/123456789/1328>.
- Townsend P. (1987) Deprivation. *J Soc Policy*; 16: 125-46.

Capitolo 11

- Bedi T., Coudouel A., Simler K. eds. (2016) *More than a pretty picture. using poverty maps to design better policies and interventions*. Washington DC: The World Bank.
- Behrens K. and Robert-Nicoud F. (2014) Survival of the Fittest in Cities: Urbanisation and Inequality. *The economic journal*, 124: 182-217.
- Boulant J., Brezzi M., and Veneri P. (2016) *Income levels and inequality in metropolitan areas: a comparative approach in OECD countries*. OECD Regional Development Working Papers 2016/06.
- Comune M.E. (2014) Un archivio longitudinale amministrativo per la stima della povertà a livello locale. *Rivista di statistica ufficiale* n.1-2/2014: 85-115.
- Comune M.E., Ronconi S., Viviano L. (2017) The challenge of permeable borders: estimating cross-border employment between Italy and Switzerland. In "Selected papers from the 2016 Conference of European statistics stakeholders". Eurostat. pp. 50-59. ISBN: 978-92-79-73629-2.
- Comune di Roma, U.O. Statistica (2016) *Il reddito dei romani. Redditi 2013*. Roma.
- Duranton G. and Puga D. (2004) Micro-foundations of urban agglomeration economies. In Henderson, J.V. and J.F. Thisse *Handbook of Regional and Urban Economics*. Vol.4, Chapter 48: 2063-2117.
- Éupolis Lombardia (2015) *La mappa lombarda del reddito. Le dichiarazioni dei redditi 2013*. Nota statistica.
- Florida R. and Mellander C. (2016) The geography of inequality: Difference and determinants of wage and income inequality across US metros. *Regional studies*, 50(1): 79-92.
- Gregory D., Johnston R., Pratt G., Watts M. and Whatmore S. (2009) *The dictionary of human geography*. Wiley- Blackwell.
- Istat (2011) *I redditi delle famiglie con stranieri*, Statistica Report, Istat, Roma.
- Istat (2015a) *La nuova geografia dei sistemi locali*, Istat, Roma. ISBN 978-88-458-1857-8
- Istat. (2015b) *Reddito e condizioni di vita*, Statistica Report, Istat, Roma.
- Istat (2015c) *Anziani: le condizioni di salute in Italia e nell'Unione Europea*, Statistica Report, Istat, Roma.
- Istat (2016) *Il benessere equo e sostenibile in Italia*, Istat, Roma. ISBN 978-88-458-1917-9.
- Ministero dell'Economia e delle Finanze (2015). *I lavoratori frontalieri italiani in Svizzera. Carburante di qualità nel motore dell'economia elvetica, in particolare ticinese*. http://www.mef.gov.it/inevidenza/documenti/doc_mef_supporto.pdf.
- Poggi P., Verrecchia F. (2004) Il progetto AMeRICA: prime indicazioni dall'integrazione di anagrafi della popolazione e dati fiscali. In Aimetti, P., B.

- Zavanella *Qualità degli archivi amministrativi e qualità dell'informazione statistica*. Milano: Franco Angeli.
- Pratesi M., eds. (2016) *Analysis of Poverty Data by Small Area Estimation*. New York: Wiley.
- Primi A. (2015) Fonti di dati e strumenti utilizzati per analisi statistiche a supporto dei processi decisionali. Relazione presentata al Convegno: *L'informazione statistica per il territorio: strumenti e buone pratiche*, Udine, 23 marzo 2015.
- Quintano C., Castellano R., and Punzo G. (2007) Estimating Poverty in the Italian Provinces using Small Area Estimation Models. *Metodološki zvezki*, 4(1): 37-70.

Capitolo 12

- European Commission, (2017) Gender Equality, sito web (al 20 gennaio 2017): http://ec.europa.eu/newsroom/just/item-detail.cfm?item_id=52696.
- Marcucci M. e Vangelisti M.I. (2013) L'evoluzione della normativa di genere in Italia e in Europa in: *Questioni di economia e finanza*, n. 188, Banca d'Italia, Roma. <http://www.bancaditalia.it/pubblicazioni/qef/2013-0188/index.html>.
- Istat (2017) Il mercato del lavoro: la lenta ripresa e le disparità nei gruppi sociali. In rapporto annuale 2017, Istat, Roma. http://www.istat.it/it/files/2017/05/RA2017_cap4.pdf.
- World Economic Forum (2016) The Global Gender Gap Report 2016. Web site: <http://reports.weforum.org/global-gender-gap-report-2016/>.
- Istat (2014) 9° Censimento dell'industria e dei servizi e Censimento delle istituzioni non profit Principali risultati e processo di rilevazione. LOMBARDIA. Istat, Roma.

Capitolo 13

- Ballabio S., Comune M.E., Ferrazza D., Verrecchia F., Vitalini A., Viviano L. (2016b) Le lenti del territorio per la qualità delle statistiche da fonte amministrativa. Convegno MILEs2016 (Milano - Impresa, Lavoro e Società 2016). Presso l'Università degli Studi di Milano Bicocca, 5 ottobre 2016. Poster.
- Ballabio S., Comune M.E., Ferrazza D., Verrecchia F., Vitalini A., Viviano L. (2016c) Chi sono i NEET? Un'analisi a partire dai dati ArchIMEDe Neeting. Convegno nazionale sui NEET. Istituto G. Toniolo di Studi Superiori, Fondazione Cariplo, Università Cattolica del Sacro Cuore Milano, Milano, 3-4 novembre 2016. Poster.
- Ballabio S., Comune M.E., Ferrazza D. (2016) Il contributo dei dati amministrativi per la conoscenza del mercato del lavoro. In: *Convegno Mutamenti sociali e differenze territoriali. Nord-sud o città-campagna?* Presso l'Aula Magna Università degli studi di Milano Bicocca. Milano, 26 ottobre 2016. Slides.
- Ballabio S., Carra A., Casacci S., Ferrazza D., Verrecchia F., Vitalini A., Viviano L. (2017) Quasi NEET? Un'analisi a partire dai dati ArchIMEDe. Neeting 2, Convegno internazionale sui NEET, Istituto G. Toniolo di Studi Superiori, Fondazione Cariplo, Università Cattolica del Sacro Cuore Milano, Centro Congressi Fondazione Cariplo, Milano, 1 dicembre 2017. Poster.
- Ferrazza D., Ballabio S., Comune M.E., Verrecchia F., Vitalini A., Viviano L. (2017) Chi sono i NEET? Un'analisi a partire dai dati ArchIMEDe. *Vita e Pensiero*, Quaderni, 6 – Rapporto Giovani: Una generazione in panchina da Neet a risorsa per il Paese. A cura di Sara Alfieri, Emiliano Sironi. pp. 58-68 ISBN Ebook (formato PDF): 978-88-343-3324-2

Istat (2014) La povertà in Italia: Anno 2013, Statistiche Report, Istat. Roma.

Capitolo 14

Della Queva S., De Francesco D., Nicosia M., Stoppiello S. (2016) “I sistemi locali del volontariato organizzato. Una nuova geografia” in Guidi R., Fonovic K., Cappadozzi T. (a cura di) *Volontari e attività volontarie in Italia. Antecedenti, impatti, esplorazioni*, Il Mulino, Bologna.

Hotelling C. S. (1973) “Resilience and Stability of Ecological System”. *Annual Review of Ecology and Systematics*, 4, 1-23.

Istat (2015a) *La nuova geografia dei sistemi locali*. ISBN 978-88-458-1857-8

Istat (2015b) *Rapporto annuale 2015 – La situazione del Paese*. Cap. 2 Luoghi, Città, Territori: struttura e dinamiche di sviluppo.

Istat (2015c) 8milaCensus. L'indice di vulnerabilità sociale e materiale, Istat, Roma.

http://ottomilacensus.istat.it/fileadmin/download/Indice_di_vulnerabilit%C3%A0_sociale_e_materiale.pdf

Martin R., & Sunley P. (2015) “On the Notion of Regional Economic Resilience: Conceptualization and Explanation”. *Journal of Economic Geography* (15), 1-42.

Conclusioni e prospettive future

Bianchino A. e Ricci M. (2017) Verso un sistema integrato di indicatori per le politiche locali. In: I Comuni verso l'uso statistico degli archivi amministrativi e dei sistemi di integrazione delle fonti. Dati e indicatori per le politiche del territorio. USCI - ISTAT, Matera 9 giugno 2017. Disponibile online <https://www.slideshare.net/slideistat/a-bianchino-m-ricci-verso-un-sistema-integrato-di-indicatori-per-le-politiche-locali>.

AUTORI

- Simona Ballabio*, Ufficio territoriale per la Lombardia, Dipartimento per la raccolta dati e lo sviluppo di metodi e tecnologie per la produzione e diffusione dell'informazione statistica, Istat.
- Maria Elena Comune*, già dell'Ufficio territoriale per la Lombardia, Dipartimento per la raccolta dati e lo sviluppo di metodi e tecnologie per la produzione e diffusione dell'informazione statistica, Istat. Oggi presso il Comune di Brescia, Settore informatica, innovazione e statistica.
- Monica Carbonara*, Ufficio territoriale per le Marche, l'Abruzzo e la Puglia, Dipartimento per la raccolta dati e lo sviluppo di metodi e tecnologie per la produzione e diffusione dell'informazione statistica, Istat.
- Salvatore Cariello*, Ufficio territoriale per la Campania e la Basilicata, Dipartimento per la raccolta dati e lo sviluppo di metodi e tecnologie per la produzione e diffusione dell'informazione statistica, Istat.
- Arianna Carra*, Ufficio territoriale per la Lombardia, Dipartimento per la raccolta dati e lo sviluppo di metodi e tecnologie per la produzione e diffusione dell'informazione statistica, Istat.
- Assunta Lisa Carulli*, Ufficio territoriale per le Marche, l'Abruzzo e la Puglia, Dipartimento per la raccolta dati e lo sviluppo di metodi e tecnologie per la produzione e diffusione dell'informazione statistica, Istat.
- Sara Casacci*, Servizio Disegno dei processi e supporto al sistema dei registri, Direzione centrale per la metodologia e il disegno dei processi statistici, Dipartimento per la raccolta dati e lo sviluppo di metodi e tecnologie per la produzione e diffusione dell'informazione statistica, Istat.
- Paola Maddalena Chiodini*, Dipartimento di statistica e metodi quantitativi, Università degli studi di Milano-Bicocca.
- Stefania Della Queva*, Servizio statistiche strutturali sulle imprese, istituzioni pubbliche e non-profit, Direzione centrale per le statistiche economiche, Dipartimento per la produzione statistica, Istat.
- Dario Ercolani*, Servizio Disegno dei processi e supporto al sistema dei registri, Direzione centrale per la metodologia e il disegno dei processi statistici, Dipartimento per la raccolta dati e lo sviluppo di metodi e tecnologie per la produzione e diffusione dell'informazione statistica, Istat.
- Daniela Ferrazza*, Ufficio territoriale per la Lombardia, Dipartimento per la raccolta dati e lo sviluppo di metodi e tecnologie per la produzione e diffusione dell'informazione statistica, Istat.
- Silvia Figini*, Dipartimento di Scienze Politiche e Sociali, Università degli studi di Pavia.
- Marianna Azzurra Filandri*, Dipartimento di Culture, Politica e Società, Università degli studi di Torino.
- Giuseppe Garofalo*, già del Servizio Disegno dei processi e supporto al sistema dei registri, Direzione centrale per la metodologia e il disegno dei processi statistici,

Dipartimento per la raccolta dati e lo sviluppo di metodi e tecnologie per la produzione e diffusione dell'informazione statistica, Istat.

Antonio Lentini, già di PoliS Lombardia - Istituto regionale per il supporto alle politiche della Lombardia.

Giancarlo Manzi, Dipartimento di Economia, Management e Metodi Quantitativi, Università degli studi di Milano.

Matteo Mazziotta, Servizio Censimento della popolazione e integrazione delle indagini sociali, Direzione centrale per le statistiche sociali e il censimento della popolazione, Dipartimento per la produzione statistica, Istat.

Manuela Nicosia, Servizio statistiche strutturali sulle imprese, istituzioni pubbliche e non-profit, Direzione centrale per le statistiche economiche, Dipartimento per la produzione statistica, Istat.

Lorella Sicuro, Ufficio territoriale per le Marche, l'Abruzzo e la Puglia, Dipartimento per la raccolta dati e lo sviluppo di metodi e tecnologie per la produzione e diffusione dell'informazione statistica, Istat.

Sabrina Stoppiello, Servizio statistiche strutturali sulle imprese, istituzioni pubbliche e non-profit, Direzione centrale per le statistiche economiche, Dipartimento per la produzione statistica, Istat.

Raffaella Succi, Ufficio territoriale per il Piemonte, la Valle D'Aosta e la Liguria, Dipartimento per la raccolta dati e lo sviluppo di metodi e tecnologie per la produzione e diffusione dell'informazione statistica, Istat.

Federico Rappelli, PoliS Lombardia - Istituto regionale per il supporto alle politiche della Lombardia.

Vincenzo Ricciari, già di Éupolis Lombardia - Istituto superiore per la ricerca, la statistica e la formazione.

Marco Trentini, Comune di Brescia, Settore informatica, innovazione e statistica.

Domenico Tucci, Ufficio territoriale per le Marche, l'Abruzzo e la Puglia, Dipartimento per la raccolta dati e lo sviluppo di metodi e tecnologie per la produzione e diffusione dell'informazione statistica, Istat.

Francesca Vannoni, Ufficio territoriale per il Piemonte, la Valle D'Aosta e la Liguria, Dipartimento per la raccolta dati e lo sviluppo di metodi e tecnologie per la produzione e diffusione dell'informazione statistica, Istat.

Alessandro Venturi, Dipartimento di Scienze Politiche e Sociali, Università degli studi di Pavia.

Flavio Verrecchia, Ufficio territoriale per la Lombardia, Dipartimento per la raccolta dati e lo sviluppo di metodi e tecnologie per la produzione e diffusione dell'informazione statistica, Istat.

Alberto Vitalini, Ufficio territoriale per la Lombardia, Dipartimento per la raccolta dati e lo sviluppo di metodi e tecnologie per la produzione e diffusione dell'informazione statistica, Istat.

Lorena Carmen Maria Viviano, Ufficio territoriale per la Lombardia, Dipartimento per la raccolta dati e lo sviluppo di metodi e tecnologie per la produzione e diffusione dell'informazione statistica, Istat.

RINGRAZIAMENTI

Uno speciale ringraziamento va ai colleghi dell'Istat che con la loro attività di ricerca hanno contribuito al progetto SPoT che ha portato alla realizzazione del presente volume. Si ringraziano anche i Dirigenti dei diversi uffici coinvolti sia territoriali¹ sia centrali² per aver sostenuto il progetto soprattutto in termini di risorse umane. In particolare, si ringrazia il Dott. Saverio Gazzelloni – Direttore della Direzione centrale per la raccolta dati dell'Istat – per aver “ospitato” il progetto tra le linee di attività della sua Direzione.

Si ringraziano anche i colleghi del Sistema statistico nazionale (Sistan) e nello specifico i ricercatori del Comune di Brescia e di PoliS Lombardia³ che hanno consentito di inserire nel volume esperienze degli enti locali.

Il progetto SPoT è stato anche l'occasione di approfondimento sul tema delle statistiche territoriali, come è avvenuto nella giornata di studio MILEs2016, e di confronto tra varie Istituzioni (p.es. Istat, enti locali e università). Da queste esperienze e sinergie sono nate collaborazioni con professori e ricercatori dell'Università degli Studi di Pavia, dell'Università degli Studi di Milano-Bicocca, dell'Università degli Studi di Torino e, in particolare, con il Dipartimento di Economia, Management e Metodi Quantitativi dell'Università degli Studi di Milano che ha anche finanziato la pubblicazione del volume.

Si ringrazia anche *Sas Institute* per gli strumenti messi a disposizione nella giornata di studio organizzata con PoliS Lombardia che ha visto la sperimentazione di nuove applicazioni “*in memory*” per la rappresentazione di fenomeni analitici complessi.

Un sentito ringraziamento per le critiche, i consigli, i suggerimenti e le proposte di revisioni che hanno migliorato il testo va al Dott. Giuseppe Garofalo e alla Dott.ssa Roberta Vivio (per il Capitolo I) e al Prof. Luigi Riva (per il Capitolo VII). Infine, ma non in ultimo, si ringraziano i revisori anonimi⁴ e il Dott. Franco Lorenzini per i preziosi suggerimenti e le revisioni proposte relativamente ad una prima versione del testo.

¹ Sede della Liguria; Sede della Lombardia; Sede del Piemonte e della Val d'Aosta; Sede dell'Abruzzo; Sede del Molise; Sede della Basilicata; Sede della Puglia.

² Servizio Disegno dei processi e supporto al sistema dei registri della Direzione centrale per la metodologia e il disegno dei processi statistici; il Servizio Censimento della popolazione e integrazione delle indagini sociali della Direzione centrale per le statistiche sociali e il censimento della popolazione; il Servizio statistiche strutturali sulle imprese, istituzioni pubbliche e non-profit della Direzione centrale per le statistiche economiche.

³ Istituto regionale per il supporto alle politiche della Lombardia.

⁴ Revisori del Comitato scientifico-redazionale del Dipartimento per la raccolta dati e lo sviluppo di metodi e tecnologie per la produzione e diffusione dell'informazione statistica.

Parte prima - Introduzione



Parte seconda - Introduzione



Parte terza - Introduzione



Conclusioni



Il volume è il frutto di una ricca attività di ricerca realizzata grazie alle conoscenze, esperienze e competenze maturate in diversi contesti territoriali e ambiti di ricerca e di produzione. È il prodotto della sinergia tra l'accademia, l'Istituto Nazionale di Statistica (Istat) e gli enti locali realizzatasi all'interno del progetto SPoT – dati e metodologie per lo sviluppo delle Statistiche per le Politiche Territoriali – avviato nel 2016 nell'ambito delle linee di attività della Direzione centrale per la raccolta dati dell'Istat.

L'obiettivo principale del progetto è la sperimentazione, con dati di fonte amministrativa, di approcci informativi di supporto alle politiche di livello locale. La prospettiva è quella di ampliare l'utilizzo dei dati amministrativi per la produzione di nuova informazione ad un livello di granularità sub-provinciale. In linea con una delle principali missioni dell'Istat, si propone così un rinnovamento dell'attenzione sulla necessità di sviluppare un'approfondita conoscenza della realtà ambientale, economica e sociale dell'Italia ai diversi livelli territoriali e di favorire i processi decisionali di tutti i soggetti della società. L'ottica sperimentale se, da un lato, può non rispettare i criteri necessari all'introduzione di nuove metodologie, alla sua traduzione in soluzioni tecnologiche e organizzative, al rispetto dei requisiti di qualità e delle regole di armonizzazione, dall'altro lato, ha un potenziale elevatissimo in quanto colma lacune conoscitive in maniera tempestiva e con periodicità frequente, dà impulso allo sviluppo di nuove analisi e di nuovi indicatori, garantisce un valido sostegno conoscitivo alle politiche.

Il volume si articola in quattro parti: *Prima parte*, dati, metadati e qualità delle fonti amministrative. *Seconda parte*, approcci gestionali e informativi, tecniche per la costruzione di indicatori e metodi grafici rivolti alla diffusione di informazioni territoriali di supporto alle politiche dei governi locali. *Terza parte*, applicazioni Sistan, dove risulta più forte il nesso tra pianificazione, programmazione e gestione, nell'ottica di soddisfare la necessità di statistiche a supporto di specifici processi decisionali pubblici. *Quarta parte*, casi di studio Istat sviluppati nell'ambito di tematiche connesse alla Deprivazione e al Benessere Equo e Sostenibile nei Territori (DeBEST).

Flavio Verrecchia – *project manager* dell'iniziativa SPoT – è primo tecnologo presso l'Istituto Nazionale di Statistica dove ha ricoperto numerosi incarichi di responsabilità di rilievo nazionale. Collabora con diversi atenei italiani. È professore a contratto di statistica presso il Corso di Laurea in Scienze del Lavoro, dell'Amministrazione e del Management dell'Università degli Studi di Milano.