SINAPPSI
CONNESSIONI TRA RICERCA E POLITICHE PUBBLICHE
Rivista quadrimestrale dell'Istituto nazionale per l'analisi delle politiche pubbliche

L'utilizzo delle banche dati amministrative in Italia per informare le politiche

Editoriale
L'utilizzo delle banche dati amministrative in Italia per informare le politiche
Fabio Berton, Roberto Leombruni

I dati amministrativi: le regole del GDPR per l'uso a fini statistici e di ricerca
Chiara di Somma

Note su avventure e sventure dei dati amministrativi, con riferimento all'esperienza di Veneto Lavoro
Bruno Anastasia

Perché essere interoperabili
Alessio Buonocore, Eugenio Concetti, Giampiero D'Alessandro, Adriana Fidanza, Stefania Pantoni, Roberta Radini

Open data e orientamento delle politiche pubbliche nella politica di coesione
Giorgia Marinuzzi, Walter Tortorella

VisitINPS e l'uso dei dati amministrativi per la conoscenza e la valutazione delle politiche pubbliche
Maria De Paola, Monica Paiella, Gianfranco Santoro

Istruzione, salari ed eterogeneità delle imprese: evidenze da dati employer-employee linked
Irene Brunetti, Michelangelo Filippi, Valentina Gualtieri, Andrea Ricci

La mobilità regionale degli studenti universitari
Andrea Priulla, Dalit Contini, Sonia Della Monica, Noemi Ferrara

Integrazione dei brevetti e delle imprese italiane
Sergio Salamone

Dagli aggregati ai microdati
Clara Andresciani

La profilazione delle persone disponibili al lavoro in Piemonte
Monica Bosia, Marianna Carlone, Elisa Floredan, Francesca Prever

TREMOD-Imprese: un modello di microsimulazione
Martina Bazzoli, Sonia Marzadro, Annalisa Tassi, Carlo Fiorio

Saggi

Ripensare la leadership: verso un nuovo modello di professionalità 5.0
Anna Grimaldi, Alfonso Monsurrò

L'istituzionalizzazione della partecipazione pubblica in Italia
Vittoria Laino

Anno XVI n. 2/2026
INAPP Editore

SINAPPSI

CONNESSIONI TRA RICERCA E POLITICHE PUBBLICHE

Rivista scientifica quadrimestrale di analisi dei fenomeni sociali e valutazione delle politiche che riguardano l'occupazione e lo sviluppo delle risorse umane, la formazione, l'inclusione e la protezione sociale

Direttore editoriale
Natale Forlani

Direttore responsabile
Pierangela Ghezzo

Comitato scientifico
Ignazia Maria Bartholini; Leonardo Becchetti; Fabio Berton; Giancarlo Blangiardo; Mario Ciampi; Giuseppe Croce; Francesco Dini; Francesca Gelli; Cristiano Gori; Marcella Milana; Catia Nicodemo; Mariella Nocenzi; Giampiero Proia; Giancarlo Rovati; Antonello Scialdone; Maria Enrica Virgillito

Comitato di redazione
Pierangela Ghezzo,
Valeria Cioccolo, Monia De Angelis,
Ernestina Greco, Paola Piras, Aurelia Tirelli
sinappsi@inapp.gov.it

INAPP Editore
00198 Roma - Corso d'Italia, 33
Tel. +39 06 854471
www.inapp.gov.it

Iscrizione al Tribunale di Roma
n. 420/2010 del 21/10/2010
© 2026 Inapp

Quest'opera è rilasciata sotto i termini della licenza Creative Commons Attribuzione - Non commerciale Condividi allo stesso modo 4.0. Italia License. (http://creativecommons.org/licenses/by-nc-sa/4.0/)



ISBN: 978-88-543-0390-4
ISSN: 2532-8549
E-ISSN: 2611-6332

Pubblicato ad agosto 2026
Finito di stampare
nel mese di settembre 2026
da Ediguida s.r.l
84013 Cava De' Tirreni (SA)

Le opinioni espresse dagli autori non impegnano la responsabilità di SINAPPSI, né quella dell'Inapp.

L'Istituto nazionale per l'analisi delle politiche pubbliche (INAPP) è un Ente pubblico di ricerca che si occupa di analisi, monitoraggio e valutazione delle politiche del lavoro, delle politiche dell'istruzione e della formazione, delle politiche sociali e, in generale, di tutte le politiche economiche che hanno effetti sul mercato del lavoro. Nato il 1° dicembre 2016 a seguito della trasformazione dell'Isfol e vigilato dal Ministero del Lavoro e delle politiche sociali, l'Ente ha un ruolo strategico – stabilito dal decreto legislativo 14 settembre 2015, n. 150 – nel nuovo sistema di governance delle politiche sociali e del lavoro del Paese. L'Inapp fa parte del Sistema statistico nazionale (SISTAN) e collabora con le istituzioni europee. È Organismo Intermedio del Programma nazionale Giovani, donne e lavoro 2021-2027 del FSE+, delegato dall'Autorità di Gestione all'attuazione di specifiche azioni (Piano Inapp 2023-2029), ed è Agenzia nazionale del Programma comunitario Erasmus+ per l'ambito istruzione e formazione professionale. È l'Ente nazionale all'interno del consorzio europeo ERIC-ESS che conduce l'indagine European Social Survey. L'attività dell'Inapp si rivolge a una vasta comunità di stakeholder: ricercatori, accademici, mondo della pratica e policymaker, organizzazioni della società civile, giornalisti, utilizzatori di dati, cittadinanza in generale.

Presidente
Natale Forlani

Direttore generale
Loriano Bigi

Sommario

L'utilizzo delle banche dati amministrative in Italia per informare le politiche

a cura di Fabio Berton e Roberto Leombruni

4 Editoriale
L'utilizzo delle banche dati amministrative in Italia per informare le politiche
Fabio Berton, Roberto Leombruni

24 I dati amministrativi: le regole del GDPR per l'uso a fini statistici e di ricerca
Chiara di Somma

34 Note su avventure e sventure dei dati amministrativi, con riferimento all'esperienza di Veneto Lavoro
Bruno Anastasia

49 Perché essere interoperabili
Ontologie, vocabolari controllati e schemi di dati del Catalogo nazionale dati
Alessio Buonocore, Eugenio Concetti, Giampiero D'Alessandro, Adriana Fidanza, Stefania Pantoni, Roberta Radini

65 Open data e orientamento delle politiche pubbliche nella politica di coesione
Dall'accountability al policy learning
Giorgia Marinuzzi, Walter Tortorella

77 VisitINPS e l'uso dei dati amministrativi per la conoscenza e la valutazione delle politiche pubbliche
Maria De Paola, Monica Paiella, Gianfranco Santoro

91 Istruzione, salari ed eterogeneità delle imprese: evidenze da dati employer-employee linked
Irene Brunetti, Michelangelo Filippi, Valentina Gualtieri, Andrea Ricci

106 La mobilità regionale degli studenti universitari
Opportunità e prime evidenze da una sperimentazione di linkage tra Anagrafe nazionale studenti e banca dati WHIP
Andrea Priulla, Dalit Contini, Sonia Della Monica, Noemi Ferrara

122 **Integrazione dei brevetti e delle imprese italiane**
Sergio Salamone

135 **Dagli aggregati ai microdati**
Spesa sanitaria privata e incentivi fiscali in Italia
Clara Andresciani

145 **La profilazione delle persone disponibili al lavoro in Piemonte**
Il ruolo dei dati amministrativi e dei dati territoriali locali
Monica Bosia, Marianna Carlone, Elisa Floredan, Francesca Prever

161 **TREMOD-Imprese: un modello di microsimulazione**
Martina Bazzoli, Sonia Marzadro, Annalisa Tassi, Carlo Fiorio

Saggi

174 **Ripensare la leadership: verso un nuovo modello di professionalità 5.0**
Anna Grimaldi, Alfonso Monsurrò

189 **L'istituzionalizzazione della partecipazione pubblica in Italia**
Una proposta di periodizzazione dagli anni Ottanta
Vittoria Laino

207 **Scaffale – Rubrica di recensioni**
Lusenti N., ***Il Coraggio di contare. Storie di donne, finanza ed etica nell'Italia contemporanea,*** Il Saggiatore, 2024
(Cristiana Porcarelli)
Gramigna A., Scialdone A. (a cura di), ***Diritto o privilegio? Fabbisogni abitativi e politiche locali per la casa***, Egea, 2026
(Enrico Puccini)
Riccioni I., ***Per una sociologia del confine. Spazio e identità nella società contemporanea,*** Carocci, 2025
(Simona Testana)

213 Per proporre un articolo

215 Norme bibliografiche

Editoriale

L'utilizzo delle banche dati amministrative in Italia per informare le politiche

Fabio Berton
European Commission, Joint Research Centre

Roberto Leombruni
Università di Torino

1. Perché i dati amministrativi?

Il fascicolo 2/2026 della rivista *Sinappsi* è il primo di due numeri speciali dedicati allo stato dell'arte sull'utilizzo delle banche dati amministrative a scopo di ricerca in Italia. L'obiettivo è quello di evidenziare ed esemplificare le caratteristiche che li rendono particolarmente adatti all'analisi statistica e alla valutazione delle politiche pubbliche, come testimonia non solo la crescente domanda di accesso a queste banche dati da parte della comunità scientifica, ma anche la crescente attenzione politica a livello italiano ed europeo.

Per dati amministrativi intendiamo "*data collected by public entities for their service activities*" (Crato e Paruolo 2019, v); più precisamente, intendiamo i microdati raccolti da tali enti nell'espletamento delle loro funzioni, che riguardano quindi informazioni a livello del singolo individuo, famiglia, impresa ecc., senza aggregazioni. La "lunga e tortuosa" storia del loro utilizzo a scopo di ricerca scientifica in Italia (Leombruni e Ardito 2026), inizia a metà degli anni Ottanta del secolo scorso quando Bruno Contini, oggi Professore Emerito di Econometria dell'Università di Torino, e Riccardo Revelli – troppo prematuramente scomparso – ebbero l'idea di trasformare le serie dei contributi previdenziali conservate presso l'Istituto nazionale di previdenza sociale (Inps) in storie lavorative individuali, quelle che oggi siamo abituati a chiamare LEED (*Linked Employer-Employee Data*). Da allora questa storia si è evoluta molto, sia in ampiezza – perché oggi gli ambiti di studio coperti dall'informazione amministrativa sono cresciuti ben al di là di quanto contenuto negli archivi dell'Inps – sia in profondità, con nuove forme di accesso ai dati che cercano, con diversi gradi si successo, di coniugare flessibilità di utilizzo per i ricercatori con tutela della riservatezza del dato. Questo numero speciale della rivista *Sinappsi* si propone di dare conto, nel modo più esaustivo possibile, dell'attuale stato dell'arte in Italia dell'utilizzo delle fonti amministrative a scopo di analisi statistica.

Almeno due elementi spiegano la centralità di questo tema nel dibattito corrente. Da un lato, alcune caratteristiche dei microdati di fonte amministrativa li rendono particolarmente adatti all'analisi statistica: (i) poiché vengono raccolti dalle amministrazioni pubbliche nello svolgimento delle proprie attività, i dati amministrativi spesso coprono l'intera popolazione di riferimento, anziché un campione; ciò consente di analizzare con maggiore precisione gli effetti distributivi delle politiche in esame, e quindi di comprendere non solo se una politica funziona, ma anche per chi funziona. (ii) A differenza delle indagini campionarie, i dati amministrativi non dipendono dalla capacità degli individui di ricordare eventi passati né dalla loro disponibilità a riferire informazioni potenzialmente sensibili. Le informazioni elementari vengono infatti raccolte nel corso dell'attività amministrativa, riducendo così alcuni dei tipici problemi di errore di misura, anche involontario. (iii) Anche sotto il profilo economico il loro impiego è vantaggioso: i costi di raccolta sono già sostenuti per la finalità amministrativa originaria. (iv) Inoltre, la presenza di identificativi univoci permette di collegare archivi differenti e, quando possibile, di integrarli con dati provenienti da indagini di tipo survey, ampliando notevolmente la quantità di informazioni disponibili per l'analisi. (v) Infine, la raccolta continua rende questi dati disponibili pressoché in tempo reale e consente di seguire individui, famiglie o imprese per periodi molto lunghi.

Dall'altro lato – e almeno in parte come conseguenza di quanto appena descritto – esiste una crescente domanda, da parte della comunità scientifica e in alcuni casi anche delle stesse pubbliche amministrazioni, di accesso alle fonti amministrative integrate per incrementare la qualità e la quantità della ricerca. De Paola *et al.* nel loro contributo a questo numero monografico ci raccontano che il solo programma VistINPS, operativo dal 2016, ha sostenuto finora oltre 260 progetti di ricerca e la pubblicazione di circa 100 articoli nella collana dedicata. Volgendo lo sguardo fuori dall'Italia, Lima *et al.* (2026) danno conto di un analogo numero di progetti attivati attraverso il Microdata Research Laboratory della Banca del Portogallo (BPLIM), per un totale di circa 350 ricercatori e ricercatrici coinvolti. Il *Centre d'accès sécurisé aux données* (CASD) dell'Istituto nazionale di statistica in Francia ha, dal 2012 in avanti, ospitato circa 800 progetti scientifici per oltre 1.600 ricercatori attivi in tutto il mondo. Ancora, al 2024 le 7.500 basi dati disponibili presso il *Centraal Bureau voor de Statistiek* (CBS) dei Paesi Bassi avevano attratto l'interesse di circa 1.900 ricercatori e ricercatrici (Bloem *et al.* 2026).

Il risultato di questo processo combinato è una crescente attenzione per i dati amministrativi non soltanto da parte della comunità scientifica e degli enti e istituti proprietari del dato elementare, ma anche da parte del mondo politico: sono del giugno 2024, per esempio, le conclusioni del Consiglio dell'Unione europea che invitano la Commissione europea a "*enhance access to administrative data for policy evaluation and impact assessment purposes*" (Consiglio dell'Unione europea 2024, punto 13). La stessa Commissione europea, nel luglio 2026, riconosce che "*Administrative data held by public authorities are a unique source of evidence for policy making and monitoring*" e che "*The use of administrative data can be more efficient and less costly than existing alternatives*" (Commissione europea 2026). Non ci siamo pertanto sorpresi, ma ne siamo stati ugualmente molto compiaciuti, che la *call for papers* per questo numero speciale di *Sinappsi* sull'utilizzo e sull'accesso ai dati amministrativi in Italia abbia raccolto un così ampio riscontro, al punto da richiedere di spezzare la pubblicazione dei testi accettati per la pubblicazione su due fascicoli consecutivi. Insomma, parafrasando un volume edito dallo stesso Bruno Contini assieme a Ugo Trivellato (2005), "Eppur si muove".

2. Il contenuto del numero monografico

Il fascicolo inizia con un inquadramento giuridico del tema. Il contributo di **di Somma** (*I dati amministrativi: le regole del GDPR per l'uso a fini statistici e di ricerca*) analizza l'assetto normativo italiano ed europeo che disciplina il trattamento dei dati amministrativi a fini statistici e di ricerca, con particolare riferimento al Regolamento generale sulla protezione dei dati (GDPR, Regolamento UE 2016/679). L'autrice parte dalla constatazione che il patrimonio informativo custodito dalle pubbliche amministrazioni è divenuto strategico per le politiche basate sull'evidenza, ma che gran parte delle banche dati contiene informazioni personali, in alcuni casi sensibili. In tale contesto, la normativa non impedisce il loro riutilizzo, ma ne regola rigorosamente le condizioni di liceità, le garanzie da adottare e i limiti di utilizzo. Il testo ricostruisce le origini del trattamento statistico nel Sistema statistico nazionale (Sistan) e delle norme che hanno istituito il Programma statistico nazionale e imposto l'interconnessione dei sistemi informativi pubblici. Vengono delineate le disposizioni che garantiscono il segreto statistico e il divieto di "ricadute amministrative", vietando l'impiego dei dati raccolti per scopi diversi da quelli statistici. Il Codice della privacy si integra con le regole deontologiche per i trattamenti statistici e le modalità di accesso ai dati da parte dei ricercatori esterni, prevedendo l'obbligo di anonimizzazione o pseudonimizzazione dei dataset. La disciplina è infine integrata dalle successive modifiche che allineano il quadro nazionale al GDPR, introducendo derivazioni specifiche per le attività statistiche. Il GDPR, infatti, prevede deroghe specifiche ai principi di limitazione della finalità e di conservazione, nonché esenzioni dal divieto di trattare categorie particolari di dati, se l'utilizzo degli stessi ha finalità statistiche e a condizione che siano adottate misure adeguate di protezione, tra cui la pseudonimizzazione. In conclusione, l'autrice sostiene che, accanto al robusto apparato normativo, l'efficace utilizzo dei dati amministrativi richiede investimenti continui in misure tecniche (pseudonimizzazione, anonimizzazione, *privacyenhancing technologies*) e nel monitoraggio delle pratiche di accesso. Solo un bilanciamento attento tra tutela dei diritti fondamentali e necessità di ricerca, argomenta di Somma, potrà garantire che questi dati continuino a rappresentare un bene pubblico capace di migliorare l'evidence-based policy.

Definito il quadro giuridico, si prosegue con la testimonianza di una delle voci storicamente più importanti sull'uso delle fonti amministrative a scopo statistico, quella di Bruno **Anastasia**. Il suo contributo (*Note su avventure e sventure dei dati amministrativi, con riferimento all'esperienza di Veneto Lavoro*) analizza le opportunità e le criticità legate all'impiego dei dati amministrativi per la ricerca sociale ed economica, ponendo in evidenza come tali risorse siano strettamente connesse all'evoluzione tecnologica, alle trasformazioni istituzionali e alle interpretazioni della missione da parte degli attori pubblici. In una prima parte l'autore ribadisce che i dati amministrativi rappresentano un patrimonio di informazioni a basso costo, coprente l'intera popolazione e potenzialmente conformi ai principi FAIR della European Data Strategy – per la quale i dati devono essere *Findable, Accessible, Interoperable and Reusable* – e che se adeguatamente trattati, possano integrare o sostituire le tradizionali indagini campionarie, migliorando la tempestività e la granularità delle analisi a supporto delle politiche pubbliche. Tuttavia, la loro fruibilità è condizionata dai vincoli imposti dalla normativa sulla protezione dei dati personali, ed anche da 'gelosie' proprietarie delle amministrazioni. Il documento prosegue con una panoramica storica del mercato del lavoro italiano, evidenziando come, fino agli anni Ottanta, le statistiche fossero quasi esclusivamente basate sull'indagine campionaria Istat, con limitate capacità di disaggregazione territoriale e settoriale. Parallelamente, le attività amministrative – ad esempio le Comunicazioni Obbligatorie (CO) delle imprese e le dichiarazioni dei disoccupati nei Centri per l'impiego – producevano volumi di dati che, pur essendo poco valorizzati per costi e complessità operative, avrebbero potuto fornire indicatori più tempestivi e granulari. In questo contesto Anastasia ricostruisce il percorso dell'Osservatorio di Veneto Lavoro, con la sperimentazione dell'uso dei microdati provenienti dal software Netlabor e da successive versioni telematiche delle stesse Comunicazioni Obbligatorie. I progressi tecnologici, a partire dalla digitalizzazione delle fonti cartacee degli anni Novanta fino all'introduzione delle CO telematiche nel 2008, hanno consentito la costruzione di statistiche di flusso, più adatte all'analisi longitudinale rispetto alle tradizionali misurazioni di stock, permettendo un monitoraggio mensile dei flussi occupazionali, con tempi di rilascio inferiori a un mese, migliorando così la capacità di risposta a crisi come la recessione del 2008 e la pandemia da Covid-19. Vengono inoltre esposti casi di utilizzo dei dati amministrativi per analizzare contratti a termine, misure di incentivazione all'assunzione, profili dei lavoratori immigrati e percorsi di inserimento dei giovani e dei disabili. Tuttavia, la frammentazione delle infrastrutture regionali, le divergenze di qualità e l'assenza di una governance coordinata hanno ostacolato la realizzazione di basi dati stabili a livello regionale. Anastasia evidenzia infine l'esistenza di altre barriere residue: la mancanza di standard comuni tra le amministrazioni, la difficoltà di integrazione con altri archivi (ad esempio i dati fiscali) e la riluttanza delle istituzioni a condividere informazioni per timori di privacy e di valutazione delle proprie performance operative. In conclusione, Anastasia suggerisce che, affinché i dati amministrativi diventino un vero bene pubblico per la ricerca, sia necessario un investimento continuo in infrastrutture sicure, in regole di accesso trasparenti e in pratiche di pseudonimizzazione, nonché un mandato chiaro per la produzione statistica indipendente.

Con **Buonocore, Concetti, D'Alessandro, Fidanza, Pantoni e Radini** il numero monografico approfondisce una delle caratteristiche che rende particolarmente vantaggioso l'uso dei dati amministrativi per la ricerca, vale a dire la loro interoperabilità attraverso fonti diverse. Il loro contributo (*Perché essere interoperabili. Ontologie, vocabolari controllati e schemi di dati del Catalogo nazionale dati*) analizza la necessità di rendere interoperabili i dati amministrativi al fine di trasformarli da risorsa operativa a bene pubblico capace di alimentare politiche basate su evidenze. Gli autori inquadrano la questione nel contesto della European Data Strategy, evidenziando come i principi FAIR costituiscano la base per un mercato unico dei dati anche a livello nazionale. Il testo prosegue descrivendo la frammentazione attuale dei sistemi informativi pubblici in Italia: su quasi 12.000 enti rispondenti al censimento delle istituzioni pubbliche, solo il 15,8% detiene una banca dati interoperabile, il 24,7% potrebbe renderla tale. Questa dispersione ostacola l'integrazione, la qualità delle analisi e l'adozione di tecnologie avanzate come l'Intelligenza artificiale. A fronte di tali criticità, gli autori propongono l'interoperabilità come requisito fondamentale, articolata in quattro livelli – giuridico, organizzativo, semantico, tecnico – secondo il quadro europeo. Nel livello semantico, l'articolo si concentra sugli strumenti messi a disposizione dal Catalogo nazionale dati (schema.gov.it), gestito da Istat, che

raccoglie ontologie, vocabolari controllati e schemi di dati. Queste risorse, identificate da URI (*Uniform Resource Identifier*) persistenti e pubblicate in formati aperti (OWL, RDF, JSONLD), consentono di standardizzare concetti quali persone, organizzazioni, luoghi e servizi, facilitando la condivisione e il riuso. L'iniziativa è integrata nella Piattaforma digitale nazionale dati (PDND) e nel principio *once only*, che permette ad un'amministrazione di condividere dati già in possesso con altre entità autorizzate, riducendo duplicazioni e costi. Il contributo ne evidenzia inoltre le implicazioni per la politica pubblica: dati interoperabili migliorano la tempestività e la precisione delle decisioni, favoriscono valutazioni di impatto più robuste, aumentano la trasparenza e la responsabilità amministrativa e aprono la strada a servizi digitali innovativi basati su IA. Tuttavia, gli autori avvertono che la piena valorizzazione richiede governance condivisa, risorse infrastrutturali, competenze tecniche e il rispetto rigoroso del GDPR. In conclusione, sostengono gli autori, la standardizzazione semantica e l'interoperabilità dei dati amministrativi rappresentano un fattore abilitante cruciale per trasformare il patrimonio informativo pubblico in un'infrastruttura di conoscenza capace di sostenere una Pubblica amministrazione più efficiente, trasparente e orientata al valore pubblico.

Con il contributo di **Marinuzzi e Tortorella** (*Open data e orientamento delle politiche pubbliche nella politica di coesione. Dall'accountability al policy learning*) si evidenzia la capacità dei dati amministrativi di essere, oltre che uno strumento di trasparenza informativa, anche una vera infrastruttura di orientamento delle politiche. Il caso di studio è offerto dagli open data nella politica di coesione europea. Marinuzzi e Tortorella osservano che, sebbene gli open data fossero tradizionalmente intesi come un '*addon*' reputazionale, nel periodo di programmazione 2021-2027 – e ancor più in quello successivo 2028-2034 – la flessibilità, il reindirizzamento delle priorità e la rapidità di attuazione sono diventati elementi strutturali della politica di coesione. In questo contesto, la questione non è più se pubblicare i dati, ma come costruire un sistema informativo aperto capace di trasformare i dati amministrativi in capacità di definire priorità, selezionare strumenti, gestire tradeoff territoriali, correggere in corso d'opera le azioni programmatiche e favorire l'apprendimento comparativo tra programmi e territori. Il caso italiano è illustrato attraverso il passaggio dagli elenchi dei beneficiari pubblicati dalle singole Autorità di Gestione fino allo sviluppo dell'iniziativa OpenCoesione, che dal 2012 mette a disposizione dataset dettagliati su programmazione e attuazione dei fondi di coesione. L'articolo ricostruisce le difficoltà dei primi cicli (2007-2013) – distribuzione eterogenea degli elenchi dei beneficiari, formati non standardizzati, assenza di licenze di riuso e scarsa qualità dei file PDF – e mostra come il Regolamento generale sui Fondi strutturali 2014-2020 abbia introdotto l'obbligo di pubblicare le informazioni in formato aperto, con licenze di riuso e metadati più ricchi. La successiva normativa sui fondi 2021-2027 ha consolidato OpenCoesione come portale unico, garantendo l'accessibilità dei dati in formato CSV, aggiornamenti periodici e la possibilità di scaricare l'intera base dati a livello di singolo progetto. Marinuzzi e Tortorella sottolineano che la mera disponibilità di dataset non genera impatto automatico. La trasparenza funzionale, o 'targeted transparency', è necessaria per rendere le informazioni comparabili e azionabili. In questo senso, gli open data supportano tre funzioni proprie delle politiche pubbliche: l'allocazione informata delle risorse, l'adattamento in itinere delle misure e l'apprendimento comparativo. Il contributo dimostra, analizzando i Fondi FESR e FSE+ 2021-2027, come la granularità dei dati di OpenCoesione consenta di rilevare pattern di spesa, identificare colli di bottiglia, valutare la distribuzione territoriale dei beneficiari e monitorare la microprogettualità, soprattutto nel FSE+. Infine, l'articolo descrive l'interoperabilità di OpenCoesione con altre banche dati quali la Banca dati delle Amministrazioni pubbliche BDAP-MEF, ReGiS e Italia Domani, attraverso identificativi univoci comuni. Tale integrazione permette di incrociare informazioni su soggetti beneficiari, localizzazione e tipologia di finanziamento, rendendo possibile un'analisi più robusta dei flussi di risorse e delle loro conseguenze sui territori. Marinuzzi e Tortorella ne concludono che OpenCoesione è una risorsa capace di guidare le decisioni politiche, facilitare la riallocazione delle risorse e supportare un apprendimento comparato a livello europeo. Tuttavia, avvertono che il pieno utilizzo di questi dati richiede ancora il rispetto di standard di metadati, la continuità della pubblicazione e l'integrazione strutturata con i cicli decisionali, elementi imprescindibili per trasformare i dati amministrativi in bene pubblico efficace.

Il contributo di **De Paola, Paiella e Santoro** (*VisitINPS e l'uso dei dati amministrativi per la conoscenza e la valutazione delle politiche pubbliche*) si focalizza su quella che è, ad oggi, probabilmente

l'esperienza più nota di accesso strutturato agli archivi amministrativi italiani a scopo di analisi statistica, ovvero quella maturata all'interno dell'Istituto nazionale della previdenza sociale (Inps) soprattutto attraverso i programmi VisitINPS e più recentemente VisitPNRR-Lavoro. Gli archivi dell'Inps includono dati contributivi, informazioni su pensioni, prestazioni di welfare e registri aziendali, e coprono circa 25 milioni di lavoratori, 1,7 milioni di imprese e 16 milioni di pensionati. Questi dataset garantiscono copertura totale della popolazione, continuità temporale e accuratezza derivante da registrazioni legali, consentendo analisi longitudinali e studi su gruppi poco numerosi. L'accesso avviene attraverso quattro canali: osservatori statistici (dati aggregati gratuiti), open data (dataset anonimizzati scaricabili), fornitura dati su richiesta (dataset personalizzati a pagamento per soggetti non Sistan) e la piattaforma VisitINPS, che fornisce microdati ad alta granularità in ambienti protetti, previa selezione mediante bandi competitivi. Il programma VisitINPS ha finora supportato più di 260 progetti, generato 94 WorkINPS Papers e organizzato otto convegni, dimostrando un forte interesse da parte della comunità scientifica. Un'estensione più recente, VisitPNRR-Lavoro, sfrutta le stesse infrastrutture per valutare le misure del Piano nazionale di ripresa e resilienza, integrando dati Inps con quelli della Ragioneria generale dello Stato. De Paola, Paiella e Santoro individuano quattro linee di sviluppo futuro: ampliamento dell'accesso, maggiore interoperabilità tra archivi, integrazione con survey mirate e diffusione di microdati anonimizzati o sintetici. Si sottolinea la necessità di potenziare le risorse amministrative e tecniche, migliorare la documentazione dei dataset e offrire percorsi formativi per ricercatori, in particolare per giovani studiosi. L'articolo conclude che, sebbene vi siano criticità operative e di governance, l'esperienza VisitINPS rappresenta, almeno in Italia, un modello per l'accesso sicuro ai dati, capace di generare evidenze empiriche utili al policy making e di promuovere una cultura dell'evidence-based policy.

Con il contributo di **Brunetti, Filippi, Gualtieri e Ricci** (*Istruzione, salari ed eterogeneità delle imprese: evidenze da dati employer-employee linked*) il potenziale degli archivi longitudinali dell'Inps, in continuità con il contributo di De Paola *et al.* sopra richiamato, viene ulteriormente evidenziato attraverso l'integrazione con le informazioni provenienti dalle Comunicazioni Obbligatorie nazionali – che Inapp prepara a scopo di analisi statistica su mandato del Ministero del Lavoro – e con dettagli qualitativi e organizzativi provenienti da una fonte di tipo survey come la RIL (Rilevazione Imprese e Lavoratori), un'indagine campionaria sulle imprese operanti in Italia che Inapp produce a partire dal 2005, arrivata nel 2025 alla settima edizione, ribadendo con ciò come l'interoperabilità delle banche dati permetta di evitare la duplicazione di informazioni quando queste possono essere più convenientemente integrate attraverso fonti diverse. Questa integrazione permette infatti di associare a ciascun lavoratore non solo la propria storia occupazionale e retributiva, ma anche le caratteristiche di governance, di relazioni industriali, di proprietà familiare e di appartenenza a reti datoriali dell'impresa in cui lavora, informazioni tipicamente assenti nelle fonti di tipo amministrativo. Il risultato è una base di conoscenza più strutturata e potente, in grado di supportare sia la ricerca accademica sia la definizione di politiche del lavoro più mirate, evidenziando il potenziale di un approccio basato su dati integrati per comprendere la relazione tra istruzione e mercato del lavoro. Per dimostrare questo potenziale, autori e autrici si chiedono come le caratteristiche strutturali delle imprese possano modellare i rendimenti salariali dell'istruzione universitaria, concentrandosi sui lavoratori laureati. Nello specifico, l'analisi intende indagare in che misura la differenza di retribuzione associata al titolo di studio sia riconducibile alle specifiche politiche salariali delle imprese, oltre che alle caratteristiche individuali dei lavoratori. I risultati principali mostrano che le politiche aziendali influenzano in modo positivo, seppur debole, l'evoluzione dei premi salariali dei laureati, essendo questi ultimi riconducibili prevalentemente alla componente legata alle eterogeneità e abilità degli occupati. Le evidenze illustrano, poi, come la performance competitiva – misurata dal fatturato per addetto – non si trasferisce automaticamente in un incremento dei premi salariali per i laureati e quando ciò accade avviene solo parzialmente e nel medio periodo. In sintesi, le politiche retributive aziendali costituiscono un canale mediante il quale il capitale umano dei lavoratori può essere (o meno) valorizzato, ma la misura in cui questo canale contribuisce a spiegare l'evoluzione dei rendimenti salariali dell'istruzione universitaria dipende in modo rilevante dall'organizzazione dei mercati interni e dalle specifiche caratteristiche delle imprese.

Rimanendo nell'alveo degli studi sull'istruzione, **Priulla, Contini, Della Monica, Ferrara** (*La mobilità*

regionale degli studenti universitari. Opportunità e prime evidenze da una sperimentazione di linkage tra Anagrafe nazionale studenti e banca dati WHIP) mirano a valutare la fattibilità e il valore aggiunto dell'integrazione fra l'Anagrafe nazionale degli studenti (ANS) e la banca dati WHIP (Work Histories Italian Panel) al fine di ricostruire in maniera longitudinale i percorsi di studio e lavoro degli individui, con particolare attenzione alla mobilità regionale di laureati e di studenti che abbandonano gli studi. L'obiettivo è dimostrare come tale integrazione possa superare le limitazioni dei dataset di tipo survey come AlmaLaurea, consentendo una più accurata analisi delle dinamiche di selezione territoriale e delle disuguaglianze regionali. L'ANS è un archivio amministrativo gestito dal Ministero dell'Università e della ricerca che registra, per tutti gli studenti iscritti alle università italiane (laureati e non), dati anagrafici, informazioni sulla scuola superiore, cronologia annuale di immatricolazione, corsi, crediti, cambi di Ateneo e data di laurea. Le sue potenzialità risiedono nella copertura totale della popolazione studentesca e nella granularità temporale che permette di monitorare percorsi accademici complessi. L'accesso ai dati è regolato da un protocollo di ricerca con il MUR, con restrizioni di utilizzo e protezione dei dati personali. WHIP, gestita dal Ministero della Salute, è una banca dati che ricostruisce le biografie lavorative e di salute di circa il 7% dei lavoratori italiani mediante il collegamento di Inps, Inail, Ministero della Salute e Comunicazioni Obbligatorie. Offre dati longitudinali dettagliati su contratti, regioni di lavoro, settore e professione, ma è limitata ai soggetti che hanno presentato una dichiarazione ISEE. L'accesso a WHIP avviene all'interno del Programma statistico nazionale, tramite l'Università di Torino che è tra i partecipanti della relativa scheda del PSN. L'abbinamento tra le fonti in questo caso è di tipo probabilistico ed è basato su sesso, data e Comune di nascita. Le indagini sulla mobilità territoriale mostrano che i laureati, soprattutto a livello magistrale, hanno una propensione al trasferimento verso il Nord-Ovest e verso aree di lavoro più qualificate, con tassi di mobilità del 30% (magistrali) e 23% (triennali). Gli studenti che abbandonano gli studi invece rimangono prevalentemente nella macroarea dell'Ateneo, con tassi di ritorno alla zona di origine superiori (circa 4550% contro 3035% per i laureati). Priulla *et al.* sottolineano che l'integrazione ANS-WHIP consente di analizzare con maggiore precisione le traiettorie studiolavoro, fornendo una piattaforma metodologica per studiare le disuguaglianze regionali e valutare politiche di istruzione e lavoro. Raccomandano tuttavia lo sviluppo di infrastrutture di *linkage* più stabili, in ossequio al principio di piena interoperabilità più volte sottolineato.

Un ulteriore caso di *linkage* tra dati amministrativi privi di chiavi univoche comuni è quello presentato da **Salamone** (*Integrazione dei brevetti e delle imprese italiane*). Il suo contributo intende dimostrare come l'integrazione dei dati brevettuali con le informazioni sulle imprese possa fornire una base empirica robusta per la progettazione, l'attuazione e la valutazione delle politiche nazionali di ricerca-sviluppo, trasferimento tecnologico e industriale. Le due fonti principali sono lo European Patent Office (EPO) e i registri statistici delle imprese di fonte Istat (l'Archivio statistico delle imprese attive ASIA e il relativo Registro delle unità locali). L'EPO fornisce i record delle domande di brevetto e delle concessioni dal 2015 al 2023, includendo informazioni su inventori, richiedenti, data di deposito, classificazione IPC (International Patent Classification) e citazioni. Questi dati sono completi, standardizzati e resi pubblici tramite il server di pubblicazione europeo, ma presentano limiti legati al fatto che sono osservate solo le invenzioni brevettate, escludendo innovazioni tacite o protette da segreto industriale. ASIA, ASIA-unità locali e Frame SBS contengono dati anagrafici, di localizzazione (ragione sociale, Comune, CAP, indirizzo) ed economici (dimensione, settore, produttività) delle imprese attive. Essi garantiscono una copertura pressoché totale del panorama aziendale privato non agricolo e permettono analisi longitudinali. L'accesso ad entrambe le fonti è regolato da procedure istituzionali: i dati EPO possono essere acquisiti dal European Publication Server, mentre i dataset Istat sono disponibili per la ricerca previa autorizzazione. I dati in questione, tuttavia, non contengono una chiave di identificazione univoca in comune. Utilizzando pertanto un *record linkage* basato su ragione sociale, Comune, CAP e indirizzo, l'autore costruisce un panel di 16.492 richiedenti, di cui 11.892 sono imprese, e ottiene una copertura del 95% delle domande di brevetto (32.133 record). Le imprese rappresentano il 72% degli applicant e l'89% delle domande, con una predominanza di imprese che presentano un unico brevetto (62,6%), mentre il 7,8% mantiene una propensione persistente alla brevettazione. La dimensione aziendale è una determinante chiave: le imprese con più di 250 dipendenti forniscono quasi la metà dei brevetti, mentre la quota delle micro-imprese con brevetti è in crescita dal 21% al 30% nel periodo analizzato.

Dal punto di vista settoriale, la maggior parte dei brevetti è concentrata nel manifatturiero, ma i settori emergenti – micro-nanotecnologie, semiconduttori, intelligenza artificiale e tecnologie ambientali – mostrano i più alti tassi di variazione, suggerendo aree di intervento prioritario per le politiche di transizione digitale e verde. Le imprese brevettanti hanno una produttività del lavoro media di circa 116.000 euro per dipendente, quasi tre volte quella delle imprese non brevettanti, indicando una forte correlazione tra attività inventiva codificata e performance economica.

L'articolo di **Andresciani** (*Dagli aggregati ai microdata. Spesa sanitaria privata e incentivi fiscali in Italia*) ci mostra invece quali criticità sorgono quando non è possibile accedere a un archivio di microdati di fonte amministrativa, e si deve ricorrere a una pluralità di fonti informative, ciascuna caratterizzata da specifici punti di forza e limiti. L'obiettivo dell'autrice è fornire una valutazione critica delle principali sfide legate alla misurazione della spesa sanitaria privata, della copertura assicurativa volontaria e delle detrazioni fiscali in Italia, evidenziando come la frammentazione delle informazioni, l'eterogeneità tra le fonti e, in alcuni casi, l'elevato livello di aggregazione dei dati possano ostacolare la capacità di monitorare l'equità e l'efficacia delle politiche di finanziamento sanitario. Le banche dati esaminate comprendono i conti nazionali dei consumi classificati secondo lo scopo (c.d., COICOP - Classification of Individual COnsumption by Purpose), il Sistema di Conti sanitari, l'indagine sui consumi delle famiglie dell'Istat e i dati amministrativi del Sistema Tessera sanitaria (STS). I conti nazionali e il Sistema dei Conti sanitari forniscono stime aggregate della spesa privata, ma differiscono per schema contabile, perimetro di riferimento e inclusione di ticket o agevolazioni fiscali, rendendo complessi i confronti tra fonti. L'indagine Istat sui consumi delle famiglie offre dati a livello nazionale e regionale, ma è soggetta a variabilità campionaria e possibile sottodichiarazione. Il STS, gestito dal Ministero dell'Economia e delle finanze, registra le transazioni sanitarie a livello individuale, ma le pubblicazioni attuali forniscono solo informazioni aggregate, senza distinguere tra spesa '*out of pocket*' (OOP), rimborsi assicurativi o ticket, e limitando l'analisi distributiva. In modo non sorprendente, i risultati principali mostrano quindi che (i) le stime della spesa sanitaria privata variano sensibilmente a seconda della fonte; (ii) la mancanza di microdati impedisce di distinguere chi sostiene la spesa, in che misura e con quali coperture assicurative, ostacolando analisi di equità territoriale e socioeconomica; (iii) le detrazioni fiscali sono riportate in forma aggregata, senza consentire di identificare se i benefici ricadono sui ticket del SSN o su spese interamente private, compromettendo la valutazione della loro progressività. Andresciani evidenzia che, se il STS fosse reso accessibile a livello micro per la ricerca con adeguate tutele di privacy, sarebbe possibile scomporre la spesa privata in pagamenti OOP, rimborsi assicurativi e ticket, e valutare la distribuzione degli incentivi fiscali per gruppi di reddito e regioni. Il rafforzamento della governance dei dati sulla spesa sanitaria e l'accesso a set più granulari appaiono pertanto prerequisiti imprescindibili per monitorare efficacemente il ruolo del finanziamento privato nei sistemi sanitari ibridi, per verificare l'effettivo 'effetto decongestione' da parte del sistema sanitario privato e per garantire che le politiche fiscali non generino effetti regressivi.

Gli ultimi due contributi di questo numero speciale di Sinappsi esemplificano infine il potenziale delle fonti amministrative anche per la cosiddetta analisi ex ante. L'obiettivo dell'articolo di **Bosia, Carlone, Floredan e Prever** (*La profilazione delle persone disponibili al lavoro in Piemonte. Il ruolo dei dati amministrativi e dei dati territoriali locali*) è valutare se l'integrazione di dati amministrativi individuali con informazioni territoriali dei Sistemi locali del Lavoro (SLL) possa migliorare la capacità predittiva dei modelli di profilazione utilizzati nei servizi per l'impiego, con particolare attenzione al caso piemontese. I dati amministrativi impiegati derivano dalla Scheda anagrafico-professionale (SAP), un fascicolo digitale che aggrega informazioni anagrafiche, formative, occupazionali e di partecipazione alle politiche attive, alimentato automaticamente dalle Comunicazioni Obbligatorie e dai Centri per l'impiego. La SAP consente di osservare l'intera popolazione di soggetti che sottoscrivono la Dichiarazione di Immediata Disponibilità (DID), garantendo copertura completa, granularità elevata e aggiornamento continuo. Per arricchire l'analisi, autori e autrici attingono al Sistema informativo dei SLL, prodotto da Istat, che combina registri delle imprese (ASIA), del lavoro (con il Registro base degli Individui, RBI) e dell'anagrafe per costruire indicatori territoriali. Queste variabili forniscono una descrizione dettagliata della struttura produttiva e demografica dei territori locali. I risultati principali mostrano che il modello replicato a livello regionale riproduce molte delle regolarità individuate dal modello nazionale, ma con

differenze marcate: l'esperienza lavorativa recente ha un impatto preventivo sulla non occupazione più forte in Piemonte, mentre l'età e il genere risultano meno determinanti. L'inclusione di variabili a livello di SLL riduce l'errore di previsione, evidenziando che un contesto con elevata quota di addetti stranieri aumenta il rischio di non reinserimento, mentre una più alta presenza di lavoratori indipendenti è associata a un minor rischio, a patto che l'autoimpiego sia strutturale. I risultati suggeriscono inoltre che l'impiego di algoritmi più flessibili può individuare più efficacemente i soggetti a più alto rischio di esclusione prolungata dal mercato del lavoro.

Arriviamo così all'ultimo contributo incluso in questo numero speciale, quello di **Bazzoli, Marzadro, Tassi e Fiorio** (*TREMOD-Imprese: un modello di microsimulazione*), il qui obiettivo è quello di presentare TREMOD-Imprese, un modello di microsimulazione fiscale che consente di valutare ex ante le politiche di fiscalità d'impresa a livello della Provincia autonoma di Trento. TREMOD-Imprese si basa su un set di dati amministrativi costituiti dalle dichiarazioni IRES, IRAP e IVA dell'anno fiscale 2019, integrati con l'Archivio Istat delle imprese attive (ASIA). I dati fiscali forniscono informazioni dettagliate su utili, valore della produzione, volume d'affari, aliquote, detrazioni e crediti d'imposta per ciascuna impresa, mentre l'anagrafica consente di ricavare dimensione (classi di addetti), settore, sede legale e altri caratteri strutturali. L'integrazione avviene mediante il codice fiscale dell'impresa, garantendo una corrispondenza univoca tra le tre dichiarazioni e l'archivio ASIA. L'accesso a tali banche dati è reso possibile dall'accordo attuativo Tremod tra l'Istituto di Statistica della Provincia di Trento (ISPAT) e la Fondazione Bruno Kessler, rientrante nel Protocollo d'intesa per l'attuazione di ricerche statistiche, e che assicura la sicurezza delle informazioni e il rispetto della riservatezza. Le simulazioni esemplificative dimostrano che la rimozione delle detrazioni IRES comporterebbe un incremento di 1,6 milioni di euro (circa < 1% del gettito IRES), mentre l'eliminazione delle aliquote ridotte porterebbe a un aumento di 5,2 milioni di euro (≈ 2%). Per l'IRAP, l'aumento dell'aliquota standard al livello nazionale (da 2,68% a 3,9%) produrrebbe un incremento di circa 33 milioni di euro (≈ 33% del gettito IRAP), mentre l'eliminazione delle deduzioni regionali genererebbe solo un aumento marginale di 1,1 milioni di euro. Le simulazioni sono statiche, non prevedono reazioni comportamentali delle imprese e sono limitate dalla mancanza di dati su crediti d'imposta specifici e su alcune variabili necessarie per scenari più complessi. Autori e autrici concludono che l'integrazione di dati fiscali amministrativi offre un valido strumento per la valutazione ex ante di politiche fiscali anche a livello subnazionale, ma evidenziano la necessità di ampliare la copertura temporale e di arricchire il set di variabili per consentire simulazioni più sofisticate e attendibili, per esempio per includere reazioni di tipo comportamentale.

3. Questioni aperte

I contributi raccolti in questo numero di *Sinappsi* mostrano con chiarezza il potenziale dei dati amministrativi per la produzione di evidenza empirica e per il miglioramento delle politiche pubbliche, ma mettono anche in evidenza alcuni ostacoli che in Italia ne limitano ancora la piena valorizzazione. Le questioni aperte non riguardano soltanto la disponibilità materiale dei dati: investono le condizioni giuridiche e organizzative dell'accesso, la capacità di integrare fonti differenti, la qualità e la documentazione degli archivi, nonché le infrastrutture necessarie a renderli effettivamente utilizzabili.

1. **Governance, coordinamento e regole comuni.** Autori e autrici evidenziano innanzitutto la frammentazione dei sistemi informativi e l'assenza, in alcuni ambiti, di una governance coordinata. L'esperienza dei dati sul mercato del lavoro descritta da Anastasia, per esempio, mostra divergenze nella qualità delle infrastrutture regionali, mancanza di standard comuni e difficoltà di costruire basi dati stabili. Analogamente, la condivisione delle informazioni può essere ostacolata dalla riluttanza delle istituzioni a rendere disponibili i propri archivi. Secondo gli autori e le autrici dei contributi pubblicati in questo fascicolo, il pieno utilizzo dei dati amministrativi richiede quindi regole di accesso trasparenti e un coordinamento capace di favorire la condivisione.
2. **Tutela dei dati personali e accesso alla ricerca.** Un secondo ostacolo riguarda la necessità di conciliare l'utilizzo dei dati per finalità statistiche e di ricerca con la protezione delle informazioni personali. Il quadro normativo descritto da di Somma consente il riutilizzo dei dati, ma lo sottopone a condizioni di liceità, garanzie e limiti, mentre l'accesso ai microdati richiede procedure e misure di protezione specifiche, quali la pseudonimizzazione, anonimizzazione, ambienti protetti e monitoraggio delle

pratiche di accesso. La questione aperta consiste dunque nel rendere accessibili i dati alla ricerca, nel quadro della disciplina sulla protezione dei dati personali, in modo più sistematico e trasparente di quanto accada oggi. Tutto ciò è possibile, come dimostrato dai numerosi contributi pervenuti per questo numero speciale e da altre esperienze italiane e in altri Paesi dell'Unione europea soggetti ai medesimi vincoli istituzionali vigenti in Italia (Berton e Paruolo 2026).

3. **Interoperabilità, linkage e standardizzazione.** La frammentazione delle banche dati rende centrale la capacità di collegare informazioni provenienti da fonti diverse. I contributi dedicati all'interoperabilità e ai record *linkage* mostrano però che questo processo non è automatico: le fonti possono non condividere chiavi identificative univoche (Priulla *et al.*; Salamone) e possono differire per struttura, definizioni e caratteristiche informative. Buonocore *et al.* sottolineano quindi l'importanza della standardizzazione semantica, dei vocabolari controllati e degli schemi di dati, oltre che di infrastrutture di *linkage* più stabili.
4. **Qualità, documentazione e granularità dei dati.** Un ulteriore problema riguarda la possibilità di disporre di informazioni sufficientemente dettagliate e documentate, vale a dire di microdati al livello della singola unità statistica. Il contributo di Andresciani sulla spesa sanitaria privata evidenzia in modo particolare i limiti derivanti dall'accesso ai soli dati aggregati: l'assenza di microdati impedisce di distinguere chi sostiene la spesa, in quale misura e con quali forme di copertura, ostacolando le analisi distributive. Altri contributi richiamano invece la necessità di migliorare la documentazione dei dataset, la qualità dei dati e la continuità delle informazioni pubblicate. La disponibilità del dato non è quindi sufficiente se la sua granularità, qualità o documentazione non consentono un uso analitico adeguato.
5. **Infrastrutture, risorse e competenze.** Il pieno sviluppo dell'analisi statistica richiede inoltre investimenti continuativi in capitale fisico (infrastrutture sicure) e umano (risorse tecniche e organizzative da parte delle istituzioni proprietarie dei dati ma anche percorsi formativi per i ricercatori). Anche l'esperienza maggiormente strutturata all'interno del panorama italiano, ovvero quella di VisitINPS (De Paola *et al.*), indica come linee di sviluppo l'ampliamento dell'accesso, una maggiore interoperabilità, la diffusione di microdati anonimizzati o sintetici e il rafforzamento della documentazione. La mera esistenza degli archivi amministrativi non è pertanto sufficiente; occorre la capacità di predisporre le condizioni tecniche e professionali necessarie per trasformarli in una risorsa effettivamente utilizzabile dalla ricerca.
6. **Continuità e integrazione dei dati nei processi di policy**. Infine, autori e autrici evidenziano che la disponibilità dei dati deve essere accompagnata da continuità nella pubblicazione e da un collegamento strutturato con i cicli decisionali. Il contributo di Bosia *et al.* mostra come l'integrazione di dati amministrativi con dati territoriali possa migliorare gli attuali processi decisionali che guidano l'erogazione delle politiche. L'esperienza di OpenCoesione (Marinuzzi e Tortorella), dall'altra parte, mostra che la semplice pubblicazione di dataset non produce automaticamente impatto: i dati devono essere comparabili, azionabili e disponibili con aggiornamenti periodici. A sua volta, ciò dipende da un mix di obblighi e incentivi all'utilizzo dei dati amministrativi a scopo di monitoraggio e valutazione che chiami in causa tutte le parti coinvolte: responsabili delle politiche, detentori dei dati grezzi, utilizzatori dei dati a scopo statistico.

Chiudiamo l'introduzione a questa prima parte di interventi dedicati al tema dell'utilizzo e dell'accesso alle fonti amministrative a scopo di ricerca in Italia chiedendoci perché, a fronte dei vantaggi che tali fonti offrono e che abbiamo descritto più in alto, l'Italia si trovi tutto sommato ancora a metà del guado, con tante esperienze di utilizzo interessanti dal punto di vista analitico, una tradizione piuttosto consolidata, ma un sistema nel suo complesso ancora troppo frammentato e poco strutturato per poterne sfruttare appieno tutte le potenzialità. Non riteniamo che queste ragioni siano da ricercare nella normativa a tutela della riservatezza dei dati personali e sensibili. Alcune esperienze in Italia ed altre all'interno dell'Unione europea dimostrano che, benché difficile, la costruzione di un sistema di utilizzo delle fonti amministrative a scopo di ricerca che coniughi rispetto delle norme e flessibilità nelle forme di accesso è tutt'altro che impossibile. Non riteniamo neanche che questa carenza sia da ricondurre a quelle che Berton e Paruolo (2026) etichettano come "false credenze" circa l'uso dei dati

amministrativi a scopo di ricerca. In primo luogo, circa il costo legato alla preparazione a scopo di ricerca di tali fonti: se da un lato è certamente vero che, a differenza dei dati che originano da survey, la fonte amministrativa richiede un investimento iniziale di pulizia, documentazione e preparazione dei dati grezzi, dall'altro lato tale costo è in larga misura fisso e non va sostenuto nuovamente ad ogni nuova release dei dati, a differenza di ciò che accade con le survey la cui struttura dei costi è dominata dalla componente variabile. In secondo luogo, circa i ritardi nella disponibilità del dato trattabile a scopo di analisi, non compatibili con i vincoli dell'azione politica, riteniamo invece che sia vero l'esatto contrario: il flusso di generazione del dato amministrativo grezzo è continuo per costruzione e la trasformazione in informazioni utilizzabili a scopo di analisi richiede, una volta che tali procedure siano state validate, poche settimane al massimo; viceversa i tempi di validazione ed esecuzione dei questionari sono almeno di alcuni mesi e seguono generalmente un calendario predeterminato difficilmente adattabile alle esigenze di conoscenza che emergono dalla politica.

Quelli che nel dibattito corrente vengono facilmente indicati come i 'soliti sospetti' – le norme sulla privacy, i costi, vincoli tecnici o di tempo – possono quindi spiegare solo in parte gli attuali ritardi che osserviamo in Italia. Ragioni più profonde si possono trovare estendendo il concetto di "gelosia del dato" introdotto da Anastasia, per fare riferimento a una più generale mancanza di volontà e di un adeguato sistema di incentivi che spinga le autorità preposte a investire nella direzione di un sistema coordinato e strutturato di accesso alle fonti amministrative a scopo di ricerca, con regole chiare e uniformi e sistemi di accesso che presentano costi sostenibili da parte degli utilizzatori del dato.

Bibliografia

Berton F., Paruolo P. (a cura di) (2026), *Data-driven learning in the EU. From administrative sources to public value*, Berlin, Springer (forthcoming)

Bloem M., Gorissen I., van den Boom R., van Hees V. (2026), CBS microdata: a user perspective, in Berton F., Paruolo P. (eds.), *Data-driven learning in the EU. From administrative sources to public value*, Berlin, Springer, Chapter 21 (forthcoming)

Commissione europea (2026), *Review of the 2021 European Pillar of Social Rights Action Plan*, Bruxelles, SWD (2026) 650 final

Consiglio dell'Unione europea (2024), *Draft Council Conclusions on the role of labour market, skills and social policies for resilient economies*, Bruxelles, 11066/24

Contini B., Trivellato U. (a cura di) (2005), *Eppur si muove. Dinamiche e persistenze nel mercato del lavoro italiano*, Bologna, il Mulino

Crato N., Paruolo P. (a cura di) (2019), *Data-driven policy impact evaluation. How to access microdata is transforming policy design*, Berlino, Springer

Fabre B., Gadouche K. (2026), Access to French administrative microdata, in Berton F., Paruolo P. (eds.), *Data-driven learning in the EU. From administrative sources to public value*, Berlin, Springer, Chapter 11 (forthcoming)

Leombruni R., Ardito C. (2026), The long and winding road. The access to administrative data for research in Italy, in Berton F., Paruolo P. (eds.), *Data-driven learning in the EU. From administrative sources to public value*, Berlin, Springer, Chapter 16 (forthcoming)

Lima F., Martins J.P., Neugebauer K., Pimentel J.F., Rodrigues S., Saraiva P., Sousa R., Zilhão M.J. (2026), Administrative data in Portugal, in Berton F., Paruolo P. (eds.), *Data-driven learning in the EU. From administrative sources to public value*, Berlin, Springer, Chapter 23 (forthcoming)

Using administrative data in Italy to inform policymaking

Fabio Berton
European Commission, Joint Research Centre

Roberto Leombruni
Università di Torino

1. Why administrative data?

Issue 2/2026 of the journal *Sinappsi* is the first of two special issues devoted to the state of the art in the use of administrative databases for research purposes in Italy. The aim is to highlight and illustrate the features that make administrative data particularly well suited to statistical analysis and public policy evaluation, as evidenced not only by the growing demand from the research community for access to these databases, but also by the increasing policy attention they are receiving at both Italian and European level.

By administrative data, we mean "*data collected by public entities for their service activities*" (Crato and Paruolo 2019, v); more precisely, we refer to the microdata collected by such entities in carrying out their functions, which therefore contain information at the level of the individual, household, firm, etc., rather than aggregated information. The '*long and winding*' history of their use for scientific research in Italy (Leombruni and Ardito 2026) began in the mid-1980s, when Bruno Contini, now Professor Emeritus of Econometrics at the University of Turin, and Riccardo Revelli – who passed far too soon – came up with the idea of transforming the social security contribution records held by the National Social Security Institute (INPS, Italy) into individual employment histories, what we now commonly refer to as LEED (Linked Employer-Employee Data). Since then, this history has evolved considerably, both in breadth – because the fields of study covered by administrative information now extend well beyond the material contained in INPS archives – and in depth, with new forms of data access seeking, with varying degrees of success, to reconcile researchers' flexibility of use with the protection of data confidentiality. This special issue of Sinappsi aims to provide as comprehensive an account as possible of the current state of the art in Italy regarding the use of administrative data sources for statistical analysis.

At least two factors explain the centrality of this issue in the current debate. On the one hand, several characteristics of administrative microdata make them particularly well suited to statistical analysis: (i) because they are collected by public administrations in the course of their activities, administrative data often cover the entire target population rather than a sample; this makes it possible to analyse the distributional effects of the policies under consideration more precisely, and therefore to understand not only whether a policy works, but also for whom it works. (ii) Unlike sample surveys, administrative data do not depend on individuals' ability to recall past events or on their willingness to report potentially sensitive information. Basic information is collected in the course of administrative activities, thereby reducing some of the typical problems associated with measurement error, including unintentional error. (iii) Their use is also advantageous from an economic perspective: the costs of data collection have already been incurred for the original administrative purpose. (iv) In addition, the presence of unique identifiers makes it possible to link different records and, where possible, to integrate them with survey data, substantially expanding the amount of information available for analysis. (v) Finally, continuous data collection makes these data available virtually in real time and makes it possible to follow individuals, households or firms over very long periods.

On the other hand – and at least partly as a consequence of what has just been described – there is growing demand, from the research community and in some cases from public administrations themselves, for access to integrated administrative data sources in order to improve the quality and

quantity of research. In their contribution to this special issue, De Paola *et al*. report that the VisitINPS programme alone, operational since 2016, has so far supported more than 260 research projects and the publication of around 100 articles in its dedicated series. Looking beyond Italy, Lima *et al.* (2026) report a similar number of projects carried out through the Microdata Research Laboratory of the Banco de Portugal (BPLIM), involving around 350 researchers. Since 2012, the *Centre d'accès sécurisé aux données* (CASD) of the French National Institute of Statistics has hosted around 800 scientific projects involving more than 1,600 active researchers from around the world. Likewise, as of 2024, the 7,500 databases available through the Centraal Bureau voor de Statistiek (CBS) in the Netherlands had attracted the interest of around 1,900 researchers (Bloem *et al*. 2026).

The combined outcome of these developments is growing attention to administrative data not only from the research community and from the institutions and agencies that hold the underlying data, but also from policymakers: for example, in June 2024, the Council of the European Union called on the European Commission to "enhance access to administrative data for policy evaluation and impact assessment purposes" (Council of the European Union 2024, point 13). In July 2026, the European Commission itself acknowledged that "Administrative data held by public authorities are a unique source of evidence for policy making and monitoring" and that "The use of administrative data can be more efficient and less costly than existing alternatives" (European Commission 2026). We were therefore not surprised, although we were equally pleased, that the call for papers for this special issue of Sinappsi on the use of and access to administrative data in Italy received such a strong response that publication of the accepted papers had to be split across two consecutive issues. In short, paraphrasing a volume edited by Bruno Contini himself together with Ugo Trivellato (2005), 'And yet it moves'.

The content of the special issue

The issue opens with a legal framework for the topic. **di Somma**'s contribution *(Administrative data: GDPR rules for statistical and research use)* analyses the Italian and European regulatory framework governing the processing of administrative data for statistical and research purposes, with particular reference to the General Data Protection Regulation (GDPR, Regulation (EU) 2016/679). The author starts from the observation that the information assets held by public administrations have become strategic for evidence-based policymaking, while a large proportion of administrative databases contain personal information, in some cases sensitive data. In this context, legislation does not prevent their reuse, but strictly regulates the conditions under which such processing is lawful, the safeguards to be adopted and the limits on use. The paper reconstructs the origins of statistical processing within the National Statistical System (Sistan) and the regulations that established the National Statistical Programme and required the interconnection of public information systems. It outlines the provisions guaranteeing statistical confidentiality and prohibiting administrative repercussions, thereby preventing data collected for statistical purposes from being used for other purposes. The Privacy Code is complemented by rules of professional conduct governing statistical processing and by arrangements for external researchers' access to data, including requirements for the anonymisation or pseudonymisation of datasets. The framework was subsequently supplemented by amendments aligning national legislation with the GDPR, introducing specific provisions for statistical activities. The GDPR, in fact, provides specific derogations from the principles of purpose limitation and storage limitation, as well as exemptions from the prohibition on processing special categories of personal data, where such processing is carried out for statistical purposes and appropriate safeguards are in place, including pseudonymisation. In conclusion, the author argues that, alongside a robust regulatory framework, effective use of administrative data requires continuous investment in technical measures (pseudonymisation, anonymisation and privacy-enhancing technologies) and in monitoring access practices. Only a careful balance between the protection of fundamental rights and research needs, di Somma argues, can ensure that these data continue to constitute a public good capable of improving evidence-based policymaking.

Once the legal framework has been established, the issue turns to the contribution of one of the historically most important voices in the use of administrative data sources for statistical purposes, Bruno **Anastasia**. His contribution *(Notes on the uses and challenges of administrative data: the*

experience of Veneto Lavoro) analyses the opportunities and critical issues associated with the use of administrative data for social and economic research, highlighting how such resources are closely linked to technological developments, institutional change and the ways in which public actors interpret their mission. In the first part, the author reiterates that administrative data constitute a low-cost information asset covering the entire population and potentially consistent with the FAIR principles of the European Data Strategy – according to which data should be Findable, Accessible, Interoperable and Reusable – and that, if properly processed, they can complement or replace traditional sample surveys, improving the timeliness and granularity of analyses supporting public policy. Their usability, however, is constrained by personal data protection legislation and also by administrations' proprietary "data ownership" concerns. The paper then provides a historical overview of the Italian labour market, showing that until the 1980s statistics were based almost exclusively on the Istat sample survey, with limited capacity for territorial and sectoral disaggregation. At the same time, administrative activities – for example, firms' Mandatory Communications (*Comunicazioni Obbligatorie*, CO) and declarations submitted by unemployed people at public employment services – generated large volumes of data which, although underused because of their costs and operational complexity, could have provided more timely and granular indicators. Against this background, Anastasia reconstructs the experience of the Veneto Lavoro Observatory, including experimentation with microdata from the Netlabor software and subsequent electronic versions of the same Mandatory Communications. Technological advances, from the digitisation of paper-based sources in the 1990s to the introduction of electronic CO in 2008, made it possible to construct flow statistics, better suited to longitudinal analysis than traditional stock measures, enabling monthly monitoring of employment flows, with release times of less than one month and thus improving the capacity to respond to crises such as the 2008 recession and the COVID-19 pandemic. The paper also presents cases in which administrative data have been used to analyse fixed-term contracts, hiring incentives, the profiles of immigrant workers, and the labour market integration pathways of young people and people with disabilities. However, fragmentation of regional infrastructures, differences in quality and the absence of coordinated governance have hindered the creation of stable regional databases. Anastasia also highlights other remaining barriers: the lack of common standards across administrations, difficulties in linking data with other archives (for example, tax data), and institutions' reluctance to share information because of concerns about privacy and the assessment of their operational performance. In conclusion, Anastasia suggests that, for administrative data to become a genuine public good for research, continuous investment is needed in secure infrastructures, transparent access rules and pseudonymisation practices, together with a clear mandate for independent statistical production.

With the contribution by **Buonocore**, **Concetti**, **D'Alessandro**, **Fidanza**, **Pantoni** and **Radini**, the special issue explores one of the features that makes the use of administrative data particularly advantageous for research, namely their interoperability across different sources. Their contribution *(Why be interoperable? Ontologies, controlled vocabularies and data schemas of the National Data Catalogue)* analyses the need to make administrative data interoperable in order to transform them from an operational resource into a public good capable of supporting evidence-based policymaking. The authors place the issue in the context of the European Data Strategy, highlighting how the FAIR principles provide the basis for a data single market also at national level. The paper goes on to describe the current fragmentation of public information systems in Italy: among almost 12,000 entities responding to the census of public institutions, only 15.8% hold an interoperable database, while 24.7% could make theirs interoperable. This fragmentation hampers integration, the quality of analysis and the adoption of advanced technologies such as artificial intelligence. In response to these challenges, the authors propose interoperability as a fundamental requirement, articulated at four levels – legal, organisational, semantic and technical – in line with the European framework. At the semantic level, the article focuses on the tools made available through the National Data Catalogue (schema.gov.it), managed by Istat, which brings together ontologies, controlled vocabularies and data schemas. These resources, identified through persistent URIs (Uniform Resource Identifiers) and published in open formats (OWL, RDF, JSON-LD), make it possible to standardise concepts such as people, organisations, places and services, facilitating data sharing and reuse. The initiative is integrated into the National

Digital Data Platform (PDND) and the once-only principle, which allows an administration to share data already in its possession with other authorised entities, reducing duplication and costs. The contribution also highlights the implications for public policy: interoperable data improve the timeliness and accuracy of decisions, support more robust impact assessments, increase administrative transparency and accountability, and pave the way for innovative AI-based digital services. However, the authors warn that fully realising this potential requires shared governance, infrastructure resources, technical expertise and strict compliance with the GDPR. In conclusion, they argue that semantic standardisation and interoperability of administrative data are crucial enabling factors for transforming public-sector information assets into a knowledge infrastructure capable of supporting a more efficient, transparent public administration oriented towards public value.

The contribution by **Marinuzzi** and **Tortorella** *(Open data and the orientation of public policies in cohesion policy. From accountability to policy learning)* highlights the capacity of administrative data to serve not only as a tool for information transparency but also as a genuine infrastructure for steering policy. The case study concerns open data in European cohesion policy. Marinuzzi and Tortorella observe that, although open data were traditionally viewed as a reputational 'add-on', during the 2021-2027 programming period – and even more so in the subsequent 2028-2034 period – flexibility, the redirection of priorities and speed of implementation have become structural elements of cohesion policy. In this context, the issue is no longer whether to publish data, but how to build an open information system capable of turning administrative data into the capacity to set priorities, select instruments, manage territorial trade-offs, adjust programme actions during implementation and foster comparative learning across programmes and territories. The Italian case is illustrated through the transition from beneficiary lists published by individual Managing Authorities to the development of the OpenCoesione initiative, which since 2012 has made detailed datasets on the programming and implementation of cohesion funds available. The article reconstructs the difficulties of the early programming cycles (2007-2013) – heterogeneous distribution of beneficiary lists, non-standardised formats, absence of reuse licences and poor-quality PDF files – and shows how the General Regulation on the Structural Funds for 2014-2020 introduced an obligation to publish information in open formats, with reuse licences and richer metadata. Subsequent legislation for the 2021-2027 funds consolidated OpenCoesione as a single portal, ensuring access to data in CSV format, periodic updates and the possibility of downloading the entire database at project level. Marinuzzi and Tortorella stress that the mere availability of datasets does not automatically generate impact. Functional, or "targeted", transparency is needed to make information comparable and actionable. In this sense, open data support three functions intrinsic to public policy: informed resource allocation, in-course adaptation of measures and comparative learning. Analysing the 2021-2027 ERDF and ESF+ funds, the contribution shows how the granularity of OpenCoesione data makes it possible to identify spending patterns, bottlenecks, the territorial distribution of beneficiaries and micro-project activity, especially under the ESF+. Finally, the article describes the interoperability of OpenCoesione with other databases, such as the Public Administrations Database (BDAP-MEF), ReGiS and Italia Domani, through common unique identifiers. This integration makes it possible to cross-reference information on beneficiaries, location and type of funding, enabling more robust analysis of resource flows and their consequences across territories. Marinuzzi and Tortorella conclude that OpenCoesione is a resource capable of informing policy decisions, facilitating the reallocation of resources and supporting comparative learning at European level. However, they warn that fully exploiting these data still requires compliance with metadata standards, continuity of publication and structured integration with decision-making cycles, all of which are essential for turning administrative data into an effective public good.

The contribution by **De Paola**, **Paiella** and **Santoro** *(VisitINPS and the use of administrative data for knowledge and public policy evaluation)* focuses on what is, to date, probably the best-known experience of structured access to Italian administrative archives for statistical analysis, namely the experience developed within the National Social Security Institute (INPS), especially through the VisitINPS programme and, more recently, VisitPNRR-Lavoro. INPS archives include contribution records, information on pensions, welfare benefits and business registers, and cover around 25 million workers, 1.7 million firms and 16 million pensioners. These datasets provide full population coverage, temporal

continuity and accuracy derived from legally mandated records, enabling longitudinal analyses and studies of small population groups. Access is provided through four channels: statistical observatories (free aggregated data), open data (downloadable anonymised datasets), data provision upon request (customised datasets supplied for a fee to non-Sistan entities) and the VisitINPS platform, which provides highly granular microdata in secure environments following selection through competitive calls. The VisitINPS programme has so far supported more than 260 projects, generated 94 WorkINPS Papers and organised eight conferences, demonstrating strong interest from the research community. A more recent extension, VisitPNRR-Lavoro, uses the same infrastructure to evaluate measures under the National Recovery and Resilience Plan, integrating INPS data with data from the General Accounting Department. De Paola, Paiella and Santoro identify four areas for future development: broader access, greater interoperability across archives, integration with targeted surveys, and wider availability of anonymised or synthetic microdata. They stress the need to strengthen administrative and technical resources, improve dataset documentation and provide training pathways for researchers, particularly early-career scholars. The article concludes that, although operational and governance challenges remain, the VisitINPS experience represents, at least in Italy, a model for secure data access, capable of generating empirical evidence useful for policymaking and promoting a culture of evidence-based policy.

The contribution by **Brunetti**, **Filippi**, **Gualtieri** and **Ricci** *(Education, wages and firm heterogeneity: evidence from linked employer-employee data)* further highlights the potential of INPS longitudinal archives, building on the contribution by De Paola *et al*. mentioned above, through integration with information from the national Mandatory Communications – which INAPP prepares for statistical analysis under a mandate from the Ministry of Labour – and with qualitative and organisational information from a survey source such as RIL (Survey on Enterprises and Workers), a sample survey of firms operating in Italy that INAPP has conducted since 2005 and which reached its seventh edition in 2025. This once again shows how interoperability across databases can avoid duplication of information when the same information can be more efficiently integrated through different sources. This integration makes it possible to associate each worker not only with their employment and earnings history, but also with characteristics relating to the governance, industrial relations, family ownership and membership of employers' networks of the firm in which they work, information typically absent from administrative sources. The result is a more structured and powerful knowledge base, capable of supporting both academic research and the design of more targeted labour policies, highlighting the potential of an integrated-data approach to understanding the relationship between education and the labour market. To demonstrate this potential, the authors ask how firms' structural characteristics may shape the wage returns to higher education, focusing on university graduates. Specifically, the analysis examines the extent to which the earnings differential associated with educational attainment can be attributed to firms' specific wage policies, in addition to workers' individual characteristics. The main results show that firm policies have a positive, albeit weak, effect on the evolution of graduates' wage premiums, which are predominantly attributable to heterogeneity and skill differences among workers. The evidence also shows that competitive performance – measured by turnover per employee – does not automatically translate into higher wage premiums for graduates and, when it does, the effect emerges only partially and over the medium term. In short, firms' pay policies constitute one channel through which workers' human capital may or may not be rewarded, but the extent to which this channel helps explain the evolution of returns to higher education depends substantially on the organisation of internal labour markets and on firms' specific characteristics.

Staying within the field of education research, **Priulla, Contini, Della Monica** and **Ferrara** *(Regional mobility of university students. Opportunities and preliminary evidence from a linkage procedure between the National student registry and the work histories Italian Panel)* assess the feasibility and added value of integrating the National Student Registry (ANS) with the WHIP database (Work Histories Italian Panel) in order to reconstruct individuals' study and work trajectories longitudinally, with particular attention to the regional mobility of graduates and students who drop out. The aim is to show how such integration can overcome the limitations of survey datasets such as AlmaLaurea, allowing more accurate analysis of territorial selection dynamics and regional inequalities. ANS is

an administrative archive managed by the Ministry of University and Research that records, for all students enrolled in Italian universities (graduates and non-graduates), personal data, information on upper-secondary education, annual histories of enrolment, degree programmes, credits, university transfers and graduation dates. Its potential lies in full coverage of the student population and in its temporal granularity, which makes it possible to monitor complex academic trajectories. Access to the data is governed by a research protocol with the Ministry of University and Research, with restrictions on use and protection of personal data. WHIP, managed by the Ministry of Health, is a database that reconstructs the employment and health histories of around 7% of Italian workers by linking INPS, INAIL, Ministry of Health, and Mandatory Communications data. It provides detailed longitudinal information on contracts, regions of employment, sector and occupation, but is limited to individuals who have submitted an ISEE declaration. Access to WHIP takes place within the National Statistical Programme, through the University of Turin, which is one of the participants in the relevant NSI programme entry. The linkage between the sources is probabilistic and is based on sex, date and municipality of birth. The findings on geographical mobility show that graduates, especially those holding a master's degree, tend to move towards north-western Italy and towards areas with more highly qualified employment, with mobility rates of 30% for master's graduates and 23% for bachelor's graduates. Students who drop out, by contrast, tend to remain in the macro-area of their university, with higher return-to-origin rates (around 45-50% compared with 30-35% for graduates). Priulla *et al.* stress that ANS-WHIP integration makes it possible to analyse study-to-work trajectories more accurately, providing a methodological platform for studying regional inequalities and evaluating education and labour policies. They nevertheless recommend developing more stable linkage infrastructures, in line with the principle of full interoperability repeatedly emphasised throughout the issue.

A further case of linkage between administrative data lacking common unique identifiers is presented by **Salamone** *(Linking patents and Italian firms)*. His contribution aims to show how integrating patent data with firm information can provide a robust empirical basis for the design, implementation and evaluation of national policies on research and development, technology transfer and industrial policy. The two main sources are the European Patent Office (EPO) and Istat statistical business registers (the Statistical Register of Active Enterprises, ASIA, and the related local-unit register). The EPO provides records of patent applications and grants from 2015 to 2023, including information on inventors, applicants, filing date, IPC (International Patent Classification) and citations. These data are complete, standardised and made publicly available through the European publication server, but they have limitations because they cover only patented inventions, excluding tacit innovations or innovations protected by trade secrecy. ASIA, ASIA-local units and Frame SBS contain demographic, location (company name, municipality, postcode, address) and economic data (size, sector, productivity) on active firms. They provide almost complete coverage of the non-agricultural private business population and allow longitudinal analyses. Access to both sources is governed by institutional procedures: EPO data can be obtained from the European Publication Server, while Istat datasets are available for research subject to authorisation. The sources, however, do not contain a common unique identifier. The author therefore uses record linkage based on company name, municipality, postcode and address to construct a panel of 16,492 applicants, of whom 11,892 are firms, achieving coverage of 95% of patent applications (32,133 records). Firms account for 72% of applicants and 89% of applications, with a predominance of firms filing a single patent (62.6%), while 7.8% show a persistent propensity to patent. Firm size is a key determinant: firms with more than 250 employees account for almost half of all patents, while the share of micro-firms with patents increased from 21% to 30% over the period analysed. By sector, most patents are concentrated in manufacturing, but emerging sectors – micro- and nanotechnologies, semiconductors, artificial intelligence and environmental technologies – show the highest rates of change, suggesting priority areas for digital and green transition policies. Patenting firms have average labour productivity of around €116,000 per employee, almost three times that of non-patenting firms, indicating a strong correlation between codified inventive activity and economic performance.

Andresciani's article *(From aggregate data to microdata. Private healthcare expenditure and tax incentives in Italy)* shows the challenges that arise when access to an administrative microdata archive

is not possible and a plurality of information sources must be used, each with its own strengths and limitations. The author's objective is to provide a critical assessment of the main challenges involved in measuring private healthcare expenditure, voluntary insurance coverage and tax deductions in Italy, highlighting how fragmentation of information, heterogeneity across sources and, in some cases, the high level of data aggregation may hinder the ability to monitor the equity and effectiveness of healthcare financing policies. The databases examined include national accounts for consumption classified by purpose (COICOP – Classification of Individual Consumption by Purpose), the Health Accounts System, the Istat Household Consumption Survey and administrative data from the Health Card System (STS). The national accounts and the Health Accounts System provide aggregate estimates of private expenditure, but differ in accounting framework, reference scope and treatment of co-payments or tax incentives, making comparisons across sources complex. The Istat Household Consumption Survey provides national and regional data, but is subject to sampling variability and possible under-reporting. The STS, managed by the Ministry of Economy and Finance, records healthcare transactions at individual level, but current publications provide only aggregated information, without distinguishing out-of-pocket (OOP) expenditure, insurance reimbursements or co-payments, thereby limiting distributional analysis. Unsurprisingly, the main findings therefore show that (i) estimates of private healthcare expenditure vary substantially depending on the source; (ii) the lack of microdata makes it impossible to distinguish who bears the expenditure, to what extent and with what insurance coverage, hampering analyses of territorial and socioeconomic equity; (iii) tax deductions are reported in aggregate form, making it impossible to identify whether the benefits relate to NHS co-payments or wholly private expenditure, thereby compromising assessment of their progressivity. Andresciani points out that, if the STS were made accessible at micro level for research purposes with appropriate privacy safeguards, it would be possible to break down private expenditure into OOP payments, insurance reimbursements and co-payments, and to assess the distribution of tax incentives across income groups and regions. Strengthening governance of healthcare expenditure data and access to more granular datasets therefore appear to be indispensable prerequisites for effectively monitoring the role of private financing in hybrid healthcare systems, assessing the actual "decongestion effect" of the private healthcare system, and ensuring that tax policies do not generate regressive effects.

The final two contributions to this special issue of Sinappsi illustrate the potential of administrative data sources also for so-called ex ante analysis. The article by **Bosia**, **Carlone**, **Floredan** and **Prever** *(Profiling people available for work in Piedmont. The role of administrative data and local territorial data)* assesses whether integrating individual administrative data with territorial information from Local Labour Market Systems (SLLs) can improve the predictive capacity of profiling models used by employment services, with particular reference to the Piedmont case. The administrative data used come from the Personal and Professional Record (SAP), a digital file that aggregates personal, educational, employment and participation-in-active-labour-market-policy information, automatically populated by Mandatory Communications and public employment services. The SAP covers the entire population of people who submit the Declaration of Immediate Availability (DID), ensuring complete coverage, high granularity and continuous updating. To enrich the analysis, the authors draw on the SLL information system produced by Istat, which combines business registers (ASIA), labour registers (including the Basic Individuals Register, RBI) and population registers to construct territorial indicators. These variables provide a detailed description of the productive and demographic structure of local areas. The main results show that the model replicated at regional level reproduces many of the regularities identified by the national model, but with marked differences: recent work experience has a stronger preventive effect on non-employment in Piedmont, while age and gender are less important determinants. Including SLL-level variables reduces prediction error, showing that a context with a high share of foreign workers increases the risk of failure to re-enter employment, whereas a higher presence of self-employed workers is associated with lower risk, provided that self-employment is structural. The results also suggest that the use of more flexible algorithms may identify people at higher risk of prolonged labour-market exclusion more effectively.

We thus come to the final contribution included in this special issue, by **Bazzoli, Marzadro, Tassi,** and **Fiorio** *(TREMOD-Imprese: a microsimulation model)*, whose objective is to present TREMOD-Imprese,

a fiscal microsimulation model that makes it possible to assess ex ante business taxation policies at the level of the Autonomous Province of Trento. TREMOD-Imprese is based on a set of administrative data comprising IRES, IRAP and VAT tax returns for fiscal year 2019, integrated with the Istat register of active enterprises (ASIA). The tax data provide detailed information on profits, production value, turnover, tax rates, deductions and tax credits for each firm, while the business register provides size (employee classes), sector, registered office and other structural characteristics. Integration is performed using the firm's tax identification number, ensuring a one-to-one match between the three tax returns and the ASIA register. Access to these databases is made possible by the Tremod implementing agreement between the Statistical Institute of the Autonomous Province of Trento (ISPAT) and the Bruno Kessler Foundation, under the Memorandum of Understanding for the implementation of statistical research, which ensures information security and confidentiality. The illustrative simulations show that removing IRES deductions would increase revenue by €1.6 million (approximately <1% of IRES revenue), while eliminating reduced tax rates would increase it by €5.2 million (≈2%). For IRAP, raising the standard rate to the national level (from 2.68% to 3.9%) would generate an increase of around €33 million (≈33% of IRAP revenue), whereas eliminating regional deductions would produce only a marginal increase of €1.1 million. The simulations are static, do not incorporate firms' behavioural responses, and are constrained by the lack of data on specific tax credits and on some variables needed for more complex scenarios. The authors conclude that integrating administrative tax data provides a useful tool for ex ante evaluation of tax policies also at sub-national level, but they stress the need to extend the time coverage and enrich the set of variables in order to enable more sophisticated and reliable simulations, for example by incorporating behavioural responses.

Open questions

The contributions collected in this issue of *Sinappsi* clearly demonstrate the potential of administrative data for producing empirical evidence and improving public policies, but they also highlight several obstacles that still limit their full exploitation in Italy. The open questions concern not only the physical availability of data, but also the legal and organisational conditions governing access, the ability to integrate different sources, the quality and documentation of archives, and the infrastructures needed to make them genuinely usable.

1. **Governance, coordination and common rules.** Authors highlight, first of all, the fragmentation of information systems and the absence, in some areas, of coordinated governance. The experience with labour-market data described by Anastasia, for example, shows differences in the quality of regional infrastructures, a lack of common standards and difficulties in building stable databases. Likewise, information sharing may be hindered by institutions' reluctance to make their archives available. According to the authors of the contributions published in this issue, fully exploiting administrative data therefore requires transparent access rules and coordination capable of fostering data sharing.
2. **Personal data protection and research access.** A second obstacle concerns the need to reconcile the use of data for statistical and research purposes with the protection of personal information. The regulatory framework described by di Somma allows data to be reused, but subjects such reuse to conditions of lawfulness, safeguards and limitations, while access to microdata requires specific procedures and protection measures, such as pseudonymisation, anonymisation, secure environments and monitoring of access practices. The open question is therefore how to make data accessible to research, within the framework of personal data protection rules, in a more systematic and transparent way than is currently the case. This is possible, as demonstrated by the numerous contributions submitted for this special issue and by other experiences in Italy and in other European Union countries subject to the same institutional constraints in force in Italy (Berton and Paruolo 2026).
3. **Interoperability, linkage and standardisation.** The fragmentation of databases makes the ability to link information from different sources central. The contributions devoted to interoperability and record linkage nevertheless show that this process is not automatic: sources may not share unique identifiers (Priulla *et al*.; Salamone) and may differ in structure, definitions and informational

characteristics. Buonocore *et al*. therefore stress the importance of semantic standardisation, controlled vocabularies and data schemas, as well as more stable linkage infrastructures.

4. **Data quality, documentation and granularity.** A further problem concerns the availability of sufficiently detailed and well-documented information, namely microdata at the level of the individual statistical unit. Andresciani's contribution on private healthcare expenditure highlights in particular the limitations arising from access only to aggregate data: the absence of microdata makes it impossible to distinguish who bears the expenditure, to what extent and with what forms of coverage, hampering distributional analysis. Other contributions instead point to the need to improve dataset documentation, data quality and the continuity of published information. Data availability is therefore not sufficient if its granularity, quality or documentation does not allow for adequate analytical use.
5. **Infrastructure, resources and skills.** The full development of statistical analysis also requires sustained investment in physical capital (secure infrastructures) and human capital (technical and organisational resources within data-holding institutions, as well as training opportunities for researchers). Even the most structured experience in the Italian context, namely VisitINPS (De Paola *et al*.), identifies broader access, greater interoperability, wider dissemination of anonymised or synthetic microdata and stronger documentation as areas for development. The mere existence of administrative archives is therefore not sufficient; the technical and professional capacity needed to turn them into a resource that can actually be used by researchers must also be in place.
6. **Continuity and integration of data into policy processes.** Finally, authors stress that data availability must be accompanied by continuity in publication and by a structured connection with decision-making cycles. The contribution by Bosia *et al*. shows how integrating administrative data with territorial data can improve the decision-making processes that currently guide policy delivery. The OpenCoesione experience (Marinuzzi and Tortorella), on the other hand, shows that simply publishing datasets does not automatically generate impact: data must be comparable, actionable and available with regular updates. In turn, this depends on a mix of obligations and incentives for using administrative data for monitoring and evaluation that involves all parties concerned: policymakers, holders of raw data and users of data for statistical purposes.

We conclude this introduction to the first part of the contributions devoted to the use of and access to administrative data sources for research purposes in Italy by asking why, despite the advantages offered by such sources and described above, Italy still finds itself, broadly speaking, only halfway there, with many interesting experiences from an analytical perspective and a fairly well-established tradition, but with an overall system that remains too fragmented and insufficiently structured to exploit its full potential. We do not believe that the reasons lie in regulations protecting personal and sensitive data. Some experiences in Italy and elsewhere within the European Union show that, although difficult, building a system for using administrative data sources for research purposes that combines compliance with the rules and flexibility in access arrangements is far from impossible. Nor do we believe that this shortcoming can be attributed to what Berton and Paruolo (2026) call "false beliefs" about the use of administrative data for research. First, as regards the cost of preparing such sources for research purposes: while it is certainly true that, unlike data originating from surveys, administrative sources require an initial investment in cleaning, documenting and preparing raw data, this cost is largely fixed and does not have to be incurred again for every new data release, unlike surveys, whose cost structure is dominated by the variable component. Second, as regards delays in the availability of data that can be processed for analytical purposes, which are supposedly incompatible with the constraints of policymaking, we believe the opposite is true: the generation of raw administrative data is continuous by construction, and transforming them into information usable for analysis requires, once the relevant procedures have been validated, at most a few weeks; by contrast, the validation and fieldwork timelines for questionnaires are at least several months and generally follow a predetermined schedule that is difficult to adapt to emerging policy information needs.

The factors that are readily identified in the current debate as the 'usual suspects' – privacy rules, costs, and technical or time constraints – can therefore explain only part of the delays currently

observed in Italy. Deeper reasons may be found by extending the concept of 'data jealousy' introduced by Anastasia to refer to a more general lack of willingness and an inadequate incentive system to encourage the authorities responsible to invest in a coordinated and structured system for research access to administrative data sources, with clear and uniform rules and access arrangements whose costs are sustainable for data users.

References

Berton F., Paruolo P. (a cura di) (2026), *Data-driven learning in the EU. From administrative sources to public value*, Berlin, Springer (forthcoming)

Bloem M., Gorissen I., van den Boom R., van Hees V. (2026), CBS microdata: a user perspective, in Berton F., Paruolo P. (eds.), *Data-driven learning in the EU. From administrative sources to public value*, Berlin, Springer, Chapter 21 (forthcoming)

Commissione europea (2026), *Review of the 2021 European Pillar of Social Rights Action Plan*, Bruxelles, SWD (2026) 650 final

Consiglio dell'Unione europea (2024), *Draft Council Conclusions on the role of labour market, skills and social policies for resilient economies*, Bruxelles, 11066/24

Contini B., Trivellato U. (a cura di) (2005), *Eppur si muove. Dinamiche e persistenze nel mercato del lavoro italiano*, Bologna, il Mulino

Crato N., Paruolo P. (a cura di) (2019), *Data-driven policy impact evaluation. How to access microdata is transforming policy design*, Berlino, Springer

Fabre B., Gadouche K. (2026), Access to French administrative microdata, in Berton F., Paruolo P. (eds.), *Data-driven learning in the EU. From administrative sources to public value*, Berlin, Springer, Chapter 11 (forthcoming)

Leombruni R., Ardito C. (2026), The long and winding road. The access to administrative data for research in Italy, in Berton F., Paruolo P. (eds.), *Data-driven learning in the EU. From administrative sources to public value*, Berlin, Springer, Chapter 16 (forthcoming)

Lima F., Martins J.P., Neugebauer K., Pimentel J.F., Rodrigues S., Saraiva P., Sousa R., Zilhão M.J. (2026), Administrative data in Portugal, in Berton F., Paruolo P. (eds.), *Data-driven learning in the EU. From administrative sources to public value*, Berlin, Springer, Chapter 23 (forthcoming)

I dati amministrativi: le regole del GDPR per l'uso a fini statistici e di ricerca

Chiara di Somma
Garante per la protezione dei dati personali

Il valore del patrimonio informativo delle pubbliche amministrazioni per fini statistici e di ricerca, ivi incluse analisi previsionali, è aumentato in maniera esponenziale con l'evoluzione delle tecniche di elaborazione e studio dei dati accrescendo proporzionalmente la sua funzione strategica per le c.d. *Evidence-Based Policy*. Buona parte delle banche dati amministrative sono popolate di dati di carattere personale, spesso anche particolarmente sensibili. Questa circostanza, lungi dal costituire un ostacolo al loro riutilizzo, impone purtuttavia il rispetto di specifiche regole e condizioni. Senza pretese di esaustività, il presente contributo intende offrire una panoramica degli aspetti rilevanti della normativa in materia di protezione dei dati personali trattati per il perseguimento degli scopi di statistica, in particolare nell'ambito del Sistema statistico nazionale.

Administrative data: GDPR rules for statistical and research use

The value of the information assets held by public administrations for statistical and research purposes – including predictive analysis – has increased exponentially alongside the evolution of data processing and analysis techniques, proportionally enhancing its strategic function for so-called Evidence-Based Policies. A significant portion of administrative databases is populated with personal data, often of a sensitive nature. This circumstance, far from constituting an obstacle to their reuse, nevertheless necessitates compliance with specific rules and conditions. Without claiming to be exhaustive, this contribution aims to provide an overview of the relevant aspects of the regulatory framework concerning the protection of personal data processed for statistical purposes, with particular reference to the National Statistical System.

DOI: 10.53223/Sinappsi_2026-02-1

Citazione
di Somma C. (2026), I dati amministrativi: le regole del GDPR per l'uso a fini statistici e di ricerca, *Sinappsi*, XVI, n.2, pp.24-33

Parole chiave
Dati amministrativi
Protezione dei dati personali
Statistica

Keywords
Administrative data
Data Protection
Statistics

1. Il trattamento dei dati amministrativi nel Sistema statistico nazionale: le origini

Il quadro normativo, che più di ogni altro testimonia il valore delle banche dati amministrative come fonte informativa per orientare il decisore pubblico nelle scelte programmatiche, è quello della statistica ufficiale; la scienza dello Stato che usa i suoi stessi dati per una migliore comprensione dell'andamento, degli sviluppi, delle necessità della società (di Somma 2022).

Con la legge 23 agosto 1988, n. 400, il Governo veniva delegato "ad emanare, […], norme aventi valore di legge ordinaria per la riforma degli enti e degli organismi pubblici di informazione statistica in base ai seguenti principi e criteri direttivi: a) che sia attuato il sistematico collegamento e l'interconnessione di tutte le fonti pubbliche preposte alla raccolta e alla elaborazione dei dati statistici a livello centrale e locale; […]; d) che sia garantito il principio dell'imparzialità e della completezza nella raccolta, nella elaborazione e nella diffusione dei dati; e) che sia garantito l'accesso diretto da parte del Parlamento, delle regioni, di enti pubblici, di organi dello Stato, di persone giuridiche,

di associazioni e singoli cittadini ai dati elaborati con i limiti espressamente previsti dalla legge e nel rispetto dei diritti fondamentali della persona [...]".

In attuazione della delega è stato adottato il decreto legislativo 6 settembre 1989, n. 322 recante *Norme sul Sistema Statistico Nazionale e sulla riorganizzazione dell'Istituto nazionale di statistica, ai sensi dell'art. 24 della legge 23 agosto 1988, n. 400*. Un testo organico relativo al funzionamento del Sistema statistico nazionale (di seguito Sistan), ancora pienamente vigente e che ha subito specifiche modifiche, in particolare per un progressivo adeguamento alla normativa in materia di protezione dei dati personali[1].

Il decreto prevede l'interconnessione e il collegamento dei sistemi informativi delle pubbliche amministrazioni con il Sistan, consentendo agli uffici di statistica l'accesso a tutti i dati in possesso dell'amministrazione di appartenenza (art. 6, commi 2 e 3). Introducendo il Programma statistico nazionale (di seguito PSN), ossia lo strumento normativo triennale (aggiornato annualmente) che pianifica le indagini e le rilevazioni statistiche di interesse pubblico, il D.Lgs. 322 del 1989 ha imposto l'obbligo a tutte le amministrazioni, enti e organismi pubblici, nonché ai privati – se previsto nel PSN – di fornire i dati loro richiesti per le rilevazioni ivi previste (art. 7).

A tutela dei dati trattati, l'art. 9 impone il segreto statistico e il c.d. divieto di ricadute amministrative[2], in base al quale i dati raccolti per scopi statistici non possono essere utilizzatati per altri fini. Il decreto prevede poi, all'art. 10, una norma di raccordo tra il Sistan e i ricercatori che operano al di fuori di esso[3] per favorire la circolazione di questo importante patrimonio informativo anche nel contesto scientifico.

Il D.Lgs. n. 322 del 1989, in origine, si limitava ad individuare il punto di equilibrio tra le esigenze di accesso e di utilizzo di dati, da un lato, e le misure per il corretto ed efficace svolgimento dell'attività statistica dall'altro, facendo leva in particolare sull'indipendenza della ricerca, sul segreto statistico, e, quindi sulla natura anonima e aggregata dei relativi output. Le successive modifiche hanno integrato questo delicato equilibrio con norme volte ad armonizzare la disciplina statistica con quella in materia di protezione dei dati personali, facendo, peraltro, emergere all'atto pratico una massima convergenza tra i principi sui quali si fondano le due differenti discipline, come si avrà modo di illustrare nel prosieguo.

Determinante, infine, la più volte novellata previsione (di cui all'art. 6-bis del D.Lgs. n. 322 del 1989) relativa alle condizioni di liceità del trattamento dei dati personali ivi inclusi quelli c.d. sensibili svolti dai soggetti Sistan, in particolare, per la realizzazione delle rilevazioni previste nel PSN, la cui adozione richiede il preventivo parere del Garante per la protezione dei dati personali (di seguito anche solo Garante per la privacy o Autorità)[4].

1 A seguito dell'entrata in vigore della legge n. 675 del 31 dicembre 1996 – *Tutela delle persone e di altri soggetti rispetto al trattamento dei dati personali*, una prima modifica al D.Lgs. 322 del 1989 è stata apportata dal decreto legislativo 30 luglio 1999, n. 281, recante *Disposizioni in materia di trattamento dei dati personali per finalità storiche, statistiche e di ricerca scientifica*. Più di recente, il decreto-legge 28 gennaio 2019, n. 4 convertito con modificazioni dalla legge 28 marzo 2019, n. 26, ha novellato il D.Lgs. 322 del 1989 per armonizzarlo con il *Regolamento (UE) 2016/679 del Parlamento europeo e del Consiglio del 27 aprile 2016 Regolamento generale sulla protezione dei dati* (GDPR) nonché con il *Codice in materia di protezione dei dati personali,* D.Lgs. 30 giugno 2003, così come modificato dal D.Lgs. 10 agosto 2018, n. 101 (Codice della privacy).

2 L'art. 9 del D.Lgs. 322 del 1989 dispone che "I dati raccolti nell'ambito di rilevazioni statistiche comprese nel programma statistico nazionale da parte degli uffici di statistica non possono essere esternati se non in forma aggregata, in modo che non se ne possa trarre alcun riferimento relativamente a persone identificabili, e possono essere utilizzati solo per scopi statistici".

3 Cfr. art. 10 rubricato *Accesso ai dati statistici* del D.Lgs. 322 del 1989 e *infra* par. 5.

4 L'art. 6-bis (commi 1 e 1-*bis*) del D.Lgs. 322 del 1989 che dispone "1. I soggetti che fanno parte o partecipano al Sistema statistico nazionale possono raccogliere ed ulteriormente trattare i dati personali necessari per perseguire gli scopi statistici previsti dal presente decreto, dalla legge o dalla normativa comunitaria, qualora il trattamento di dati anonimi non permetta di raggiungere i medesimi scopi. 1-bis. Per i trattamenti di dati personali, compresi quelli di cui all'articolo 9 del Regolamento (UE) 2016/679 del Parlamento europeo e del Consiglio, del 27 aprile 2016, effettuati per fini statistici di interesse pubblico rilevante ai sensi dell'articolo 2-*sexies*, comma 2, lettera cc), del Codice in materia di protezione dei dati personali, di cui al decreto legislativo 30 giugno 2003, n. 196, in conformità all'articolo 108 del medesimo Codice, nel Programma statistico nazionale sono specificati i tipi di dati, le operazioni eseguibili e le misure adottate per tutelare i diritti fondamentali e le libertà degli interessati, qualora non siano individuati da una disposizione di legge o di regolamento. Il Programma statistico nazionale, adottato sentito il Garante per la protezione dei dati personali, indica le misure tecniche e organizzative idonee a garantire la liceità e la correttezza del trattamento, con particolare riguardo al principio di minimizzazione dei dati, e, per ciascun trattamento, le modalità, le categorie dei soggetti interessati, le finalità perseguite, le fonti utilizzate, le principali variabili acquisite, i tempi di conservazione e le categorie dei soggetti destinatari dei dati. Per i trattamenti dei dati personali di cui all'articolo 10 del citato regolamento (UE) 2016/679 effettuati per fini statistici di interesse pubblico rilevante ai sensi del citato articolo 2-sexies, comma 2, lettera cc), del codice di cui al decreto legislativo n. 196 del 2003 trova applicazione l'articolo 2-*octies* del medesimo codice) [...]". Per i trattamenti di dati personali relativi a condanne penali e reati, la normativa in esame rinvia quindi all'articolo 2-*octies* del Codice della privacy che in assenza di una specifica norma di legge o di regolamento, rinvia a sua volta ad un decreto del Ministero della Giustizia, ai sensi dell'art. 17, comma 3 della legge 23 agosto 1988, n. 400. Il Garante per la privacy, con il provvedimento n. 247 del 24 giugno 2021 (doc. web n. 9682603), ha espresso il parere su uno Schema di regolamento recante l'individuazione dei trattamenti di dati personali relativi a condanne penali e reati e delle relative garanzie appropriate, ai sensi dell'articolo 2-*octies*, comma 2 del Codice della privacy.

Ciò ha consentito la produzione di una copiosa giurisprudenza da parte dell'Autorità, che ha contribuito nel corso del tempo a definire il rapporto tra statistica e protezione dei dati personali, attraverso l'indicazione di specifici accorgimenti e misure per l'uso dei dati di carattere personale, anche inerenti alle particolari categorie, contenuti in banche date pubbliche (che qui definiamo 'dati amministrativi'), tenendo anche conto dell'evoluzione, da un canto, del processo tecnologico e, dell'altro, della scienza statistica stessa[5] (Longo 2017; Ponti 2023; Simonati 2023).

Tutto ciò premesso, si intende ora passare all'esame dei tratti del Regolamento generale sulla protezione dei dati, Regolamento (UE) 2016/679 (di seguito Regolamento o GDPR) e del Codice in materia di protezione dei dati personali, D.Lgs. 30 giugno 2003 (di seguito Codice della privacy) più rilevanti per il trattamento dei dati amministrativi a fini statistici, per poi soffermarsi sulle misure più adeguate per un equo bilanciamento tra protezione dei dati personali e attività statistica.

2. La statistica nel GDPR

Il Regolamento disciplina particolarmente il trattamento dei dati personali per finalità statistiche, dimostrando una sorta di '*favor*'[6] del Legislatore comunitario per tale scopo.

Esso, infatti, introduce nel settore in esame: una deroga ai principi di limitazione della finalità e della conservazione (art. 5, par. 1, lett. b) e e) GDPR)[7]; una specifica esenzione al divieto di trattare le particolari categorie di dati, (art. 9, par. 1 e 2, lett. j) GDPR); deroghe e limitazioni ai diritti spettanti agli interessati in relazione ai loro dati personali (artt. 14, par. 5 lett. b), 17, par. 2 lett. d), 21, par. 6, 89, par. 2). Infine, il capo IX, concernente le specifiche situazioni di trattamento, all'art. 89, rubricato *Garanzie e deroghe relative al trattamento a fini di archiviazione nel pubblico interesse, di ricerca scientifica o storica o a fini statistici*, impone l'applicazione di garanzie adeguate per i diritti e le libertà dell'interessato, facendo esplicito riferimento alla pseudonimizzazione dei dati.

Nella dimensione nazionale, in attuazione dell'art. 89 del GDPR, il Codice della privacy affida, in particolare al titolo VII – parte II (artt. da 97 a 110-bis), la disciplina dei trattamenti di dati personali a fini di ricerca scientifica e statistici introducendo: regole di carattere generale (artt. 97 e 99); un coacervo di norme dedicate alla ricerca scientifica e statistica (artt. 104-107 e 110-bis); una disciplina specifica per il trattamento di dati personali per scopi di statistica da parte dei soggetti Sistan (artt. 108-109).

In tale quadro, merita rilievo l'art. 105 del Codice della privacy che ribadisce che gli scopi di ricerca scientifica e statistici si presumono non incompatibili con quello della raccolta, qualunque essi siano, precisando che, viceversa, nessuno scopo ulteriore differente è ammesso per i dati raccolti per scopi di ricerca scientifica. Si tratta del c.d. "divieto di ricadute amministrative", elemento cardine anche dei principi internazionali ed europei per l'uso a fini statistici dei dati personali, secondo cui "dovrebbe essere severamente proibito" l'uso di dati raccolti a fini statistici per altri scopi, ad esempio "amministrativi, giuridici o fiscali", oppure per "condurre verifiche nei confronti delle unità statistiche"[8].

5 *Ex multis* si vedano i provvedimenti n. 315 del 16 settembre 2021 (doc. web n. 9717477); n. 434 del 16 dicembre 2021 (doc. web n. 9738899); n. 237 del 30 giugno 2022 (doc. web n. 9794929).

6 Sul punto occorre precisare come tale '*favor*' non possa considerarsi frutto di una decisione arbitraria ma della precisa consapevolezza che l'attività statistica – espressione di un diritto costituzionalmente tutelato (artt. 9 e 33 della Costituzione) – necessita di regole precipue per un equo bilanciamento con il diritto alla protezione dei dati personali, che al contempo non la renda inefficace o ne ostacoli indebitamente l'esercizio.

7 Merita segnalarsi in particolare che in base al "principio di limitazione della finalità", i dati sono "raccolti per finalità determinate, esplicite e legittime, e successivamente trattati in modo che non sia incompatibile con tali finalità; un ulteriore trattamento dei dati personali a fini di archiviazione nel pubblico interesse, di ricerca scientifica o storica o a fini statistici non è, conformemente all'articolo 89, paragrafo 1, considerato incompatibile con le finalità iniziali" e, che, in base al "principio di limitazione della conservazione", i dati sono "conservati in una forma che consenta l'identificazione degli interessati per un arco di tempo non superiore al conseguimento delle finalità per le quali sono trattati; i dati personali possono essere conservati per periodi più lunghi a condizione che siano trattati esclusivamente a fini di archiviazione nel pubblico interesse, di ricerca scientifica o storica o a fini statistici, conformemente all'articolo 89, paragrafo 1, fatta salva l'attuazione di misure tecniche e organizzative adeguate richieste dal presente regolamento a tutela dei diritti e delle libertà dell'interessato" (art. 5, par. 1, lett. b) e) GDPR; cfr. anche art. 99 del Codice della privacy) cfr. Gruppo articolo 29, *Opinion 03/2013 on purpose limitation,* del 2 aprile 2013; Provv. del Garante per la privacy del n 406 del 1° novembre 2021 (doc. web n. 9731827).

8 Cfr. il Considerando n. 27 del Regolamento CE N. 2009/223 sulle statistiche europee e, a livello internazionale, l'art. 4 della Raccomandazione del Consiglio d'Europa n. R (97) relativa alla protezione dei personali raccolti e trattati per scopi statistici. Il Garante per la privacy si è espresso numerosissime volte in riferimento a questo principio: cfr. *ex multis*: Provv. 271 del 9 maggio 2018 n. (doc. web n. 9001732); n. 270 del 17 dicembre 2020 (doc. web n. 9523274); n. 10 del 23 gennaio 2020 (doc. web. n. 9261093); n. 434 del 16 dicembre 2021 (doc. web n. 9738899).

3. Le condizioni di liceità del trattamento

La raccolta e il successivo trattamento di dati amministrativi per fini statistici da parte dei soggetti che fanno parte o partecipano al Sistan sono ammessi alle condizioni di cui all'art. 6, par. 1, lett. c) ed e) parr. 2 e 3 del Regolamento[9]. Qualora il trattamento sia destinato a coinvolgere dati c.d. 'sensibili', rileva invece la deroga al divieto dell'uso di tali informazioni, di cui alla lett. g), del par. 2, dell'art. 9 del GDPR (Bombardelli 2022; Califano 2017; Caravà 2019; Carullo 2024; Franca 2023; Francario 2022; Garofoli e Ferrari 2008).

In attuazione dell'art. 6, par. 3 del Regolamento, l'art. 2-quater del Codice della privacy ha conferito all'Autorità il compito di promuovere, nell'osservanza del principio di rappresentatività, l'adozione di regole deontologiche per i trattamenti previsti, tra agli altri, al capo IX del Regolamento (Bifulco 2021).

Rilevano quindi le *Regole deontologiche per trattamenti a fini statistici o di ricerca scientifica effettuati nell'ambito del Sistema statistico nazionale*, corrispondenti all'allegato A4 al Codice della privacy, che costituiscono condizione essenziale di liceità e correttezza del trattamento, nella misura in cui la loro violazione può comportare l'applicazione di una sanzione amministrativa, ai sensi dell'art. 166 del Codice della privacy[10].

4. L'art. 89, par. 1 del Regolamento e le garanzie adeguate per i diritti e le libertà dell'interessato

Quanto fin ora rappresentato dimostra come il Regolamento e per esso, laddove necessario (Baroncelli 2008), l'ordinamento nazionale, individuino ampie possibilità per l'uso dei dati amministrativi a fini statistici, fornendo un sistema di regole articolato e complesso che non si esaurisce in formali previsioni normative ma che richiede anzi obbligatoriamente altresì l'implementazione di misure adeguate sia alla tutela dei diritti e delle libertà fondamentali degli interessati sia a garantire il rispetto degli obblighi derivanti dal Regolamento. Ci si riferisce in particolare al già citato art. 89 del Regolamento[11].

Si tratta di una norma cogente che, lungi dall'avere un valore meramente indicativo, è accompagnata di uno specifico presidio sanzionatorio. L'art. 83, par. 5, lett. d) del GDPR, infatti, dispone l'applicazione di una sanzione amministrativa per la violazione di "qualsiasi obbligo ai sensi delle legislazioni degli Stati membri adottate a norma del capo IX" (nell'ordinamento interno gli artt. 97 e ss. del Codice della privacy).

L'art. 89 da una parte esorta, dall'altra sembra proprio imporre l'uso di misure di pseudonimizzazione ossia di quel trattamento a seguito del quale "i dati personali non possano più essere attribuiti a un interessato specifico senza l'utilizzo di informazioni aggiuntive, a condizione che tali informazioni aggiuntive siano conservate separatamente e soggette a misure tecniche e organizzative intese a garantire che tali dati personali non siano attribuiti a una persona fisica identificata o identificabile" (art. 4, n. 5 GDPR). In altri termini, la pseudonimizzazione sostituisce parte dei dati con uno o più codici pseudonimi[12] (Basdekis *et al.* 2023).

La reversibilità dei codici pseudonimi (che determina la natura identificativa del dato pseudonimizzato, *Cons.* 26 GDPR) è assicurata da informazioni aggiuntive, il cui trattamento richiede quindi (al pari dei dati personali) l'implementazione di specifiche misure tecniche e organizzative adeguate a garantirne la sicurezza,

9 Nell'ordinamento interno le disposizioni di cui agli artt. 6, par.1, lett. c) e e) e art. 9, par. 2, lett. g) del GDPR hanno trovato ulteriore specificazione, rispettivamente, negli artt. 2- *ter* e 2-*sexies* del Codice della privacy, secondo i quali la base giuridica del trattamento deve risiedere in una norma dell'Unione o in una norma di legge, regolamento o atto amministrativo generale dell'ordinamento interno (tra queste quindi il PSN); sono inoltre richieste ulteriori garanzie per i dati c.d. sensibili (art. 9 GDPR). Cfr. Garante europeo per la protezione dei dati (di seguito EDPS) *Assessing the necessity of measures that limit the fundamental right to the protection of personal data: A Toolkit,* 11 aprile 2017*; Guidelines on assessing the proportionality of measures that limit the fundamental rights to privacy and to the protection of personal data,* 19 dicembre 2019*; Guidance for co-legislators on key elements of legislative Proposals,* 7 maggio 2025).

10 Si segnalano altresì le *Regole deontologiche per trattamenti a fini statistici o di ricerca scientifica*, di cui all'allegato A5 del Codice. Con il provvedimento n. 133 del 15 aprile 2021 (doc. web n. 9582086). L'Autorità ha adottato la deliberazione di promovimento delle nuove *Regole deontologiche per trattamenti a fini statistici o di ricerca scientifica effettuati nell'ambito del Sistema statistico nazionale*, n. 98 doc. web n. 9107633 e con provvedimento n. 298 del 4 maggio 2024 (doc. web n. 10016146) la deliberazione di promovimento delle *Regole deontologiche per trattamenti a fini statistici o di ricerca scientifica* ai sensi degli artt. 2-quater e 106 del Codice.

11 Cfr. EDPS (European Data Protection Supervisor), *Study on the appropriate safeguards under Article 89(1) GDPR for the processing of personal data for scientific research*, *Final Report* dell' 8 febbraio 2019.

12 Cfr. Comitato europeo per la protezione dei dati (EBPB) *Guidelines 01/2025 on Pseudonymisation Adottate il 16 Gennaio 2025*; Garante per la privacy, Provvedimenti n. 10 del 23 gennaio 2020 (doc. web 9261093); n. 315, 16 settembre 2021 (doc. web n. 9717477); n. 674 del 13 novembre 2024, (doc. web n. 10090499).

tra le quali, in primo luogo, la segregazione rispetto alle informazioni pseudonimizzate[13].

La pseudonimizzazione va intesa, non solo, come una misura per ridurre i rischi correlati al trattamento dei dati personali (es. violazione della riservatezza, alterazione dei dati, *function creep*) ma è anche, laddove sapientemente impiegata congiuntamente ad altre misure tecniche e organizzative, quale utile strumento per il rispetto degli obblighi di protezione dei dati personali in generale (es. effettiva applicazione dei principi di liceità, limitazione della finalità, minimizzazione dei dati, limitazione della conservazione, obblighi di *privacy by design* e *by default*[14] misure per il trasferimento dei dati all'estero o a soggetti terzi, cfr. *Cons.* 28 GDPR). Non sarebbe, quindi, corretto fare affidamento, per l'applicazione dei principi del Regolamento e in particolare di quello di integrità e riservatezza di cui all'art. 5, par. 1 lett. f), sulla sola pseudonimizzazione dei dati, vieppiù in caso di trattamento delle particolari categorie (Stalla-Bourdillon 2025).

Si comprende tuttavia, in termini generali, la particolare adeguatezza di questa misura nel contesto della statistica. Tale scienza, infatti, è volta ad analizzare, investigare e studiare fenomeni collettivi, la loro evoluzione, la loro incidenza, le loro cause, ma è per sua natura scarsamente interessata all'identità dei singoli soggetti coinvolti in tali fenomeni. Sono piuttosto gli attributi dei singoli, spesso anche (doverosamente) generalizzati, gli elementi (*rectius* i dati) di interesse della statistica. In linea generale, infatti, pur necessitandosi di informazioni relative ad un singolo individuo, risulta eccedente e sproporzionato il trattamento anche di quelle che ne consentono l'identificazione diretta. Ciò ferma restando la possibilità di risalire all'identità del singolo laddove necessario (ad esempio a fini di monitoraggio o controlli di qualità e di copertura); possibilità che la pseudonimizzazione dei dati fa salva.

Occorre considerare, inoltre, come l'attività statistica si sia evoluta nel corso del tempo, passando da un sistema incentrato prevalentemente sulle rilevazioni dirette a uno che predilige la statistica c.d. da fonti amministrative organizzate[15]. In altri termini, buona parte delle statistiche pubbliche, ivi inclusa l'attività censuaria[16], si basa in misura sempre più preponderante sull'interconnessione, lo studio e lo svolgimento di analisi previsionali su dati amministrativi.

Gli enti Sistan utilizzano per i propri scopi molteplici fonti amministrative, riducendo così il c.d. 'fastidio statistico', tipico delle rilevazioni dirette e i costi connessi a tali operazioni, e migliorando la qualità dei dati trattati. I soggetti Sistan – e tra questi in particolare l'Istituto nazionale di statistica - Istat o gli uffici di statistica delle grandi pubbliche amministrazioni – quindi, detengono ed elaborano ulteriormente enormi quantità di dati, spesso anche sensibili, venendosi a determinare un trattamento su larga scala e ad alto rischio[17]. Si tratta di un patrimonio informativo inestimabile per la statistica, ma al tempo stesso di una fonte di elevato rischio per i diritti e le libertà fondamentali degli interessati, che richiede l'impiego di specifiche misure tecniche e organizzative di mitigazione e contenimento[18].

La pseudonimizzazione è una di queste, con l'ulteriore vantaggio che essa, nel contribuire ad assicurare il rispetto del GDPR, può operare anche a beneficio della stessa attività statistica. A titolo meramente esemplificativo è intuitivo osservare come, garantendo la riservatezza del dato, si assicuri altresì una migliore attuazione del segreto statistico o come i principi di minimizzazione ed esattezza dei dati possano altresì contribuire ad accrescere la qualità del campione statistico[19].

13 Le informazioni pseudonimizzate vengono in questo modo a costituire un 'dominio' autonomo cfr. punto 2.3 delle *Guidelines 01/2025 on Pseudonymisation* cit. e provvedimento dell'Autorità n. 10 del 23 gennaio 2020, cit.

14 Provvedimento del Garante per la privacy n. 10 del 23 gennaio 2020, cit.

15 Le statistiche da indagine (SDI) presuppongono un contatto diretto con l'interessato e, dunque, non solo la possibilità per quest'ultimo di non fornire riscontro alle domande relative ai dati sensibili o giudiziari (che sono esenti dal c.d. obbligo di risposta ai sensi dell'art. 7, comma 2, D.Lgs. n. 322/1989) ma anche che siano resi riscontri inesatti o approssimativi. Per contro, nelle statistiche da fonti amministrative organizzate (SDA) l'ente Sistan, titolare del lavoro statistico – in virtù del PSN, che funge da base giuridica del trattamento – può raccogliere dati personali, anche inerenti alle particolari categorie presumibilmente più accurati, presso altri soggetti pubblici o privati che sono tenuti a formili.

16 Cfr. art. 1, commi 227, 228, 232 della legge 27 dicembre 2017, n. 205 recante *Bilancio di previsione dello Stato per l'anno finanziario 2018 e bilancio pluriennale per il triennio 2018-2020*.

17 Cfr. art. 35 GDPR e Gruppo art. 29, *Linee guida in materia di valutazione d'impatto sulla protezione dei dati e determinazione della possibilità che il trattamento "possa presentare un rischio elevato"* ai fini del Regolamento (UE) 2016/679 adottate 4 ottobre 2017.

18 Si segnala al riguardo il Provvedimento n. 46 del 10 febbraio 2022 (doc. web n. 9751194), relativo a un *data breach* occorso ad alcuni database dell'Istat.

19 EDPS, *Guidance for Risk Management of Artificial Intelligence systems* dell'11 novembre 2025.

Assistiamo quindi ad un singolare caso in cui l'operazione di bilanciamento, necessaria per la coesistenza di attività espressione di diritti giuridicamente tutelati, piuttosto che comprimerli (facendo leva solo sulla loro naturale elasticità) ne amplifica la piena estrinsecazione. Un bilanciamento che non opera, quindi, nel 'conflitto di diritti' ma per aumentarne la convergenza.

L'art. 89 del GDPR auspica infine che il perseguimento dei fini ulteriori di statista avvenga, laddove possibile, attraverso misure che non consentano più di identificare gli interessati. Costituisce, invero, una precondizione di tutto l'impianto della normativa in materia di protezione dei dati personali l'assunto che questi vadano trattati solo se necessari[20], dovendosi, laddove possibile, preferire informazioni non personali o anonimizzate. Ciononostante le differenti tecniche di anonimizzazione[21], indispensabili in fase di diffusione dei risultati statistici, solitamente diminuiscono, dall'ottica del ricercatore, il potenziale informativo del dato, specie se comportano la perdita di granularità della informazione.

L'approccio responsabile e proattivo richiesto dal GDPR, in virtù dei principi di *accountability* e di *privacy by design* e *by default* (artt. 5, par. 2 e 25), impone tuttavia al titolare di esplorare l'utilità di talune nuove tecnologie capaci di offrire soluzioni coerenti con lo stato dell'arte tecnologica, adatte anche a valorizzare l'uso dei dati anonimizzati, ciò in particolare quando oggetto di trattamenti sono dati di natura "sensibile" (D'Acquisto e Naldi 2017).

Ci si riferisce alle c.d. *Privacy-enhancing technologies* (PET) e tra queste, in particolare, alla tecnica della sintetizzazione dei dati che offre una concreta alternativa all'uso di dati di carattere personale. Quelli sintetici, infatti, sono dati creati da strumenti di intelligenza artificiale che, partendo da dati reali, forniscono informazioni ad esse analoghe e con lo stesso valore informativo ma non identificative degli interessati (Aceto *et al.* 2024; Bassan *et al.* 2018; Boudewijn *et al.* 2023; Brusadelli *et al.* 2025; Carr *et al.* 2023; Chenna 2022; D'Acquisto 2024; El Emam *et al.* 2020). Come anticipato, la pseudonimizzazione e l'anonimizzazione sono operazioni di trattamento che incidono in maniera sostanzialmente differente sui dati personali: la prima assicurando la possibilità di risalire all'interessato, la seconda negandola.

L'anonimizzazione costituisce la *condicio sine qua non* per la diffusione dei risultati statistici e per la tutela del segreto posta a garanzia delle informazioni trattate a tale fine.

Merita brevemente quindi richiamare un importante e recente arresto della Corte di Giustizia dell'Unione europea, la sentenza del 4 settembre 2025, C-413/23, che, partendo dalle indicazioni contenute nel *Considerando* 26 del GDPR[22] ha accorciato le distanze tra tali operazioni di trattamento, esplorando la possibilità che, in determinati contesti, informazioni oggetto 'solo' di pseudonimizzazione possano comunque perdere la capacità di reindentificare gli interessati ai quali si riferiscono. Ebbene, la Corte è giunta a sostenere che in alcune circostanze, la pseudonimizzazione possa di fatto incidere sul carattere personale dei dati, a condizione che le informazioni aggiuntive siano conservate separatamente e vengano attuate misure tecniche

20 EDPS, *Assessing the necessity of measures that limit the fundamental right to the protection of personal data: A Toolkit*, cit.

21 Il dato anonimizzato, al quale non si applica la disciplina in materia di protezione dei dati personali, è tale solo se non consente l'identificazione diretta o indiretta di una persona, tenuto conto di tutti i mezzi ragionevoli (informazioni, risorse tecnologiche ed economiche, competenze, tempo) nella disponibilità di chi (titolare o altro soggetto) provi a identificare un interessato. L'anonimizzazione, deve pertanto impedire che si possa: 1. isolare una persona in un gruppo (*single-out*); 2. collegare un dato anonimizzato a dati riferibili a una persona presente in un distinto insieme di dati (*linkability*); 3. dedurre nuove informazioni riferibili a una persona da un dato anonimizzato (*inference*). Gruppo art. 29 *Parere 05/2014 sulle tecniche di anonimizzazione*, adottato il 10 aprile 2014.

22 Il *Considerando* 26 del GDPR stabilisce che "È auspicabile applicare i principi di protezione dei dati a tutte le informazioni relative a una persona fisica identificata o identificabile. I dati personali sottoposti a pseudonimizzazione, i quali potrebbero essere attribuiti a una persona fisica mediante l'utilizzo di ulteriori informazioni, dovrebbero essere considerati informazioni su una persona fisica identificabile. Per stabilire l'identificabilità di una persona è opportuno considerare tutti i mezzi, come l'individuazione, di cui il titolare del trattamento o un terzo può ragionevolmente avvalersi per identificare detta persona fisica direttamente o indirettamente. Per accertare la ragionevole probabilità di utilizzo dei mezzi per identificare la persona fisica, si dovrebbe prendere in considerazione l'insieme dei fattori obiettivi, tra cui i costi e il tempo necessario per l'identificazione, tenendo conto sia delle tecnologie disponibili al momento del trattamento, sia degli sviluppi tecnologici. I principi di protezione dei dati non dovrebbero pertanto applicarsi a informazioni anonime, vale a dire informazioni che non si riferiscono a una persona fisica identificata o identificabile o a dati personali resi sufficientemente anonimi da impedire o da non consentire più l'identificazione dell'interessato. Il presente Regolamento non si applica pertanto al trattamento di tali informazioni anonime, anche per finalità statistiche o di ricerca".

e organizzative tali da garantire in modo effettivo che i dati pseudonimizzati non siano attribuiti a una persona fisica identificata o identificabile. In altri termini, la natura identificativa dei dati pseudonimizzati dovrebbe, a mente della Corte, cessare di essere un assioma ma dipendere da una valutazione in concreto, condizionata dal contesto nel quale essi approdano e della sussistenza di mezzi che ragionevolmente possano favorire la reindentificazione degli interessati (*Cons.* 26 GDPR) (Bird&Bird 2025; He 2023; Lee 2025; Yermilov *et al.* 2023).

Ci si dovrà quindi interrogare sull'impatto di tale rinnovata interpretazione di dato pseudonimizzato e anonimizzato nell'attività statistica, specie qualora la modifica alla definizione di dato personale proposta dalla Commissione europea di Regolamento della c.d. *digital omnibus* dovesse trovare attuazione[23]. Azzardando una previsione, si ritiene che tale evoluzione potrebbe, anzi dovrebbe, avere un impatto pressoché nullo nel rigore delle valutazioni sul rischio di reindetificazioni degli interessati in fase di diffusione dei risultati statistici. Ciò, non solo per la contestuale vigenza del segreto statistico, non derogabile, ma anche per l'imprevedibilità ex ante del rischio di reidentificazione degli interessati negli innumerevoli e differenti contesti dove gli output della statistica verranno ad essere trattati a seguito della loro diffusione.

5. Dalla statistica ufficiale alla ricerca

L'uso dei 'dati amministrativi' non è appannaggio esclusivo dei soggetti Sistan. Esistono, infatti, svariati strumenti normativi affinché di tale prezioso patrimonio informativo possano godere anche i ricercatori che operano al di fuori di tale contesto, sempre ferma l'applicazione di adeguate misure (Brizzi 2021; Stefanelli 2018).

Si richiama, quindi, l'art. 10 del D.Lgs. 322 del 1989, rubricato *Accesso ai dati statistici*, che dispone che: "*1.* I dati elaborati nell'ambito delle rilevazioni statistiche comprese nel programma statistico nazionale sono patrimonio della collettività e vengono distribuiti per fini di studio e di ricerca a coloro che li richiedono secondo la disciplina del presente decreto, fermi restando i divieti di cui all'art. 9. *2.* Sono distribuite altresì, ove disponibili, su richiesta motivata e previa autorizzazione del Presidente dell'Istat, collezioni campionarie di dati elementari, resi anonimi e privi di ogni riferimento che ne permetta il collegamento con singole persone fisiche e giuridiche […]".

Al riguardo, si segnala altresì dall´art. 5-ter del decreto legislativo del 14 marzo 2013, n. 33, rubricato *Accesso per fini scientifici ai dati elementari raccolti per finalità statistiche*, in base al quale gli enti e uffici del Sistan possano consentire l´accesso, per fini scientifici, ai dati elementari, privi di ogni riferimento che permetta l´identificazione diretta delle unità statistiche, raccolti nell´ambito di trattamenti statistici di cui i medesimi soggetti siano titolari, nel rispetto di una serie di condizioni. In particolare, vengono definiti: i requisiti per l´ente di ricerca di appartenenza dei ricercatori, per il progetto di ricerca e per le cui finalità vengono richiesti i dati, le modalità di rilascio dei dati, nonché le operazioni di trattamento eseguibili, facendo divieto di effettuare trattamenti diversi da quelli previsti nel progetto di ricerca, conservare i dati elementari oltre i termini di durata del progetto, comunicare i dati a terzi e diffonderli, pena l´applicazione della sanzione di cui all´art. 162, comma 2-*bis*, del Codice della privacy.

In attuazione di tale norma, il Comitato di indirizzo e coordinamento dell'informazione statistica (Comstat), sentito il Garante per la privacy[24], ha adottato la Direttiva n. 11 del 7 novembre 2018 recante *Linee guida per l'accesso a fini scientifici ai dati elementari del Sistan*. Esse, nel disciplinare in concreto tale flusso di dati, ne assicurano la protezione, prescrivendo specifiche misure per garantirne la sicurezza e affinché i risultati della ricerca non permettano il collegamento con le unità statistiche (Patti 2019).

Merita, infine, soffermarsi sull'art. 2-ter, comma 3 del Codice privacy che consente la comunicazione di dati 'comuni' tra soggetti che trattano dati per il perseguimento di un compito di interesse pubblico (si pensi ad esempio alle università) anche laddove tale flusso non sia previsto da una norma di legge, regolamento o atto amministrativo generale a condizione che ne sia data previa notizia al Garante per la privacy almeno 10 giorni prima dell'inizio dei trattamenti (di Somma 2019).

L'ordinamento italiano dispone poi di uno specifico corpo normativo sui c.d. sistemi di sorveglianza e registri di patologia, che detta le condizioni per il trattamento dei dati sulla salute ivi raccolti anche per fini di ricerca epidemiologica, la scienza biomedica che studia la frequenza, la distribuzione e le cause di eventi legati alla salute

23 EDPB-EDPS, *Joint Opinion 2/2026 On the Proposal for a Regulation as regards the simplification of the digital legislative framework (Digital Omnibus)* del 10 febbraio 2026.

24 L'Autorità si è espressa con il Parere del 21 giugno 2018 n. 388 (doc. web 9023239).

all'interno di una popolazione e che insieme alla statistica sanitaria costituisce la base scientifica della sanità pubblica (art. 12, commi da 10 a 14 del D.L. 18 ottobre 2012, n. 179 e ss.ii.mm; D.P.C.M. 3 marzo 2017, *Cons*. 157 GDPR).

Merita in conclusione oltrepassare i confini dell'ordinamento nazionale per segnalare la grande opportunità fornita al c.d. 'uso secondario' dei dati (che include quelli statistici e di ricerca) dal Regolamento europeo n. 327 del 2025, che ha istituito lo Spazio europeo dei dati sanitari (European Health Data Space, EHDS).

Nel Regolamento EHDS, per uso secondario dei dati si intende "il trattamento dei dati sanitari elettronici per le finalità indicate al capo IV del presente Regolamento, che sono diverse dalle finalità iniziali per le quali tali dati sono stati raccolti o prodotti" (art. 2, par. 1, lett. e).

Tre le finalità (secondarie) per le quali è possibile trattare i dati sanitari elettronici, rilevano quelle "statistiche quali definite all'articolo 3, punto 1), del Regolamento (UE) n. 223/2009, come le statistiche ufficiali a livello nazionale, multinazionale e dell'Unione, relative al settore sanitario o dell'assistenza" (art. 53, par. 1, lett. c) e quelle di "ricerca scientifica nel settore sanitario o dell'assistenza che contribuisce alla sanità pubblica o alla valutazione delle tecnologie sanitarie o che garantisce elevati livelli di qualità e sicurezza dell'assistenza sanitaria, dei medicinali o dei dispositivi medici, con l'obiettivo di favorire gli utenti finali, quali i pazienti, i professionisti sanitari e gli amministratori sanitari, tra cui: i) attività di sviluppo e innovazione per prodotti o servizi; ii) attività di addestramento, prova e valutazione degli algoritmi, anche nell'ambito di dispositivi medici, dispositivi medico-diagnostici in vitro, sistemi di IA e applicazioni di sanità digitale (art. 53, par. 1, lett. e)".

Lo EHDS si inserisce perfettamente nelle maglie del GDRP nella misura in cui nell'indicare gli scopi secondari del trattamento (tra gli altri, lo si ribadisce, quello statistico e di ricerca) di determinate particolari categorie di dati e nel fornire, quindi, una solida base normativa (in attuazione degli artt. 5, par. 1 lett. a), 6 e 9 del GDPR), indica altresì le modalità con le quali concretamente i trattamenti necessari devono avere luogo, richiedendo, in particolare, la preventiva pseudonimizzazione o anonimizzazione dei dati (art. 89 GDPR e 55 e ss. EDHS) (Corso 2025; Foà 2026; Trubiani 2025)

Conclusioni

In conclusione, si è inteso in questa sede brevemente evidenziare come la centralità dei dati amministrativi per le attività di statistica e di ricerca trovi ampio riscontro in un quadro normativo nutrito, composito e articolato volto non a ostacolare il fine, ma anzi a sostenerlo, armonizzandolo al contempo con la normativa in materia di protezione dei dati personali.

Ciò non si esaurisce però con la formale osservanza delle norme, richiedendosi specifici e talvolta ingenti investimenti per l'individuazione e il costante aggiornamento di adeguate misure tecniche e organizzative e tra queste in particolare quella pseudonomizzazione dei dati[25].

Bibliografia

Aceto G., Giampaolo F., Guida C., Izzo S., Pescapè A., Piccialli F., Prezioso E. (2024), Synthetic and Privacy-Preserving Traffic Trace Generation using Generative AI Models for Training Network Intrusion Detection Systems, *Journal of Network and Computer Applications*, 229, art.103926, pp.1-16

Baroncelli S. (2008), Le fonti del diritto nell'UE dal Trattato di Roma al Trattato di Lisbona: verso un'accresciuta complessità del sistema, *Osservatoriosullefonti.it*, n.3, pp.1-46

25 Questo arresto giurisprudenziale è stato tenuto in considerazione dalla Commissione europea nella Proposta di Regolamento del Parlamento europeo e del Consiglio che modifica i Regolamenti (UE) 2016/679, (UE) 2018/1724, (UE) 2018/1725 e (UE) 2023/2854 e le Direttive 2002/58/ce, (UE) 2022/2555 e (UE) 2022/2557 per quanto riguarda la semplificazione del quadro legislativo nel settore digitale e che abroga i Regolamenti (UE) 2018/1807, (UE) 2019/1150 e (UE) 2022/868 e la Direttiva (UE) 2019/1024 (*omnibus digitale*), del 19 novembre 2025. La proposta suggerisce infatti di modificare la definizione di dato personale contenuta del GDPR, con l'aggiunta all'art, 4, n. 1 del seguente capoverso "le informazioni relative a una persona fisica non sono necessariamente dati personali per qualsiasi altra persona o entità per il solo fatto che un'altra entità può identificare tale persona fisica. Le informazioni non sono personali per una determinata entità se quest'ultima non è in grado di identificare la persona fisica cui si riferiscono le informazioni, tenendo conto dei mezzi di cui tale entità si può ragionevolmente avvalere. Tali informazioni non diventano personali per tale entità per il solo fatto che un potenziale destinatario successivo dispone di mezzi di cui si può ragionevolmente avvalere per identificare la persona fisica cui le informazioni si riferiscono".

Basdekis I., Kloukinas C., Agostinho C., Vezakis I., Pimenta A., Gallo L. Spanoudakis G. (2023), Pseudonymisation in the context of GDPR-compliant medical research, in *19th International Conference on the Design of Reliable Communication Networks (DRCN)*, Vilanova i la Geltru, Spain, 2023, pp.1-6 <DOI:10.1109/DRCN57075.2023.10108370>

Bassan S., Harel O. (2018), The Ethics in Synthetics: Statistics in the Service of Ethics and Law in Health-Related Research in Big Data from Multiple Sources, *Journal of Law and Health*, 31, n.1, pp.87-110

Bifulco R. (2021), Codici di condotta e regole deontologiche dopo il D.Lgs. n. 101/2018, in Pizzetti F. (a cura di), *Protezione dei dati personali in Italia tra GDPR e codice novellato*, Torino, Giappichelli, pp.337-350

Bird&Bird (2025), *Pseudonimizzazione e dati personali. Perché la sentenza CGUE 4 settembre2025, C-413/23 P potrebbe rappresentare un punto di svolta?,* London, Bird&Bird LLP

Bombardelli M. (2022), Dati personali (tutela dei), in Mattarella B.G., Ramajoli M. (a cura di), *Enciclopedia del diritto: i tematici*, Milano, Giuffrè, pp.351-381

Boudewijn A.T.P., Ferraris A.F., Panfilo D., Cocca V. (2023), *Privacy Measurements in Tabular Synthetic Data: State of the Art and Future Research Directions* <https://arxiv.org/abs/2311.17453>

Brizzi D. (2021), *Dati sanitari, GDPR, e Covid-19. Il caso della ricerca: tra scienza e diritto*, Milano, Giuffrè Francis Lefebvre

Brusadelli A., Ruffo D., di Somma C., Boudewijn A.T.P (2025), Ripensare l'evidenza: approccio multidisiciplinare all'uso dei dati sintetici in ambito sanitario - Parte I, *BioLaw Journal-Rivista di BioDiritto,* n.3, pp.317-340

Califano L. (2017), Il Regolamento UE 2016/679 e la costruzione di un modello uniforme di diritto europeo alla riservatezza e alla protezione dati personali, in Califano L., Colapietro C. (a cura di), *Innovazione tecnologica e valore della persona*, Napoli, Editoriale Scientifica, pp.17-44

Caravà E. (2019), Base giuridica per il trattamento di dati personali effettuato per l'esecuzione di un compito di interesse pubblico o connesso all'esercizio di pubblici poteri, in Sciaudone R., Caravà E. (a cura di), *Codice della privacy*, Pisa, Pacini Giuridica, pp.45-54.

Carr M., Wiemers E., Pollak A., McCue K. (2023), Using Synthetic Data to Estimate Earnings Dynamics Evidence from the SIPP GSF and SIPP SSB, *SSRN Electronic Journal*, pp.1-55

<DOI:10.2139/ssrn.4480922>

Carullo G. (2024), Dati personali e fini pubblici: dubbi di compatibilità europea del Codice Privacy, *Ceridap. Rivista interdisciplinare sul diritto delle amministrazioni pubbliche,* n.3, pp.37-52

Chenna S. (2022), *Application of Generative Adversarial Networks (GANs) for Generating Synthetic Data and in Cybersecurity*, *SSRN Electronic Journal*, pp.1-6 <DOI:10.2139/ssrn.4305711>

Corso S. (2025), *Autodeterminazione e dati sanitari*, Torino, Giappichelli

D'Acquisto G. (2024), Synthetic data and data protection laws*, Journal of data protection & Privacy*, 6, pp.227-236

D'Acquisto G., Naldi M. (2017), *Big data e Privacy by design*, Torino, Giappichelli

di Somma C. (2022), Ricerca scientifica storica e per finalità statistiche, in Riccio G.M., Scorza G., Belisario E. (a cura di), *GDPR e Normativa Privacy Commentario,* Milano, Ipsoa, pp.881-904

di Somma C. (2019), Trattamento in ambito pubblico e comunicazioni al Garante, *Privacy&Data Protection Technology Cybersecurity*, n. 2/2019, pp.13-16

El Emam K., Mosquera L., Hoptroff R. (2020) *Practical Synthetic Data Generation: Balancing Privacy and the Broad Availability of Data*, Sebastopol, O'Reilly Media

Foà S. (2026), L'altruismo dei dati e le sue potenzialità per l'interesse generale: dal Data governance act al diritto interno*,* in Cavallaro M.C. Fant V. (a cura di), *Ia Sanità - Territorio - Pubbliche Amministrazioni,* Napoli, Editoriale scientifica, pp.523-539

Franca S. (2023), *I dati personali Nell'amministrazione pubblica. Attività di trattamento E tutela del privato,* Napoli, Editoriale scientifica

Francario F. (2022), Protezione dei dati e pubblica amministrazione, in Pisani C., Proia G., Topo A. (a cura di), *Privacy e lavoro. La circolazione dei dati personali e i controlli nel rapporto di lavoro,* Milano, Giuffrè, pp.679-722

Garofoli R., Ferrari G. (2008), *Manuale di diritto amministrativo*, Roma, Nel Diritto editore, pp.3 e ss.

He Z. (2023), From Privacy-Enhancing to Health Data Utilisation: The Traces of Anonymisation and Pseudonymisation in EU Data Protection Law, *Digital Society*, n.2, art.17

Lee P. (2025), Synthetic Data and the Future of AI, *Cornell Law Review*, 110, n.1, pp.1-74

Longo E. (2017), I trattamenti nel settore dell'istruzione e a fini di ricerca (scientifica, storica, statistica), in Califano L., Colapietro C. (a cura di), *Innovazione tecnologica e valore della persona*, Napoli, Editoriale scientifica, pp.491-508

Yermilov O., Raheja V., Chernodub A. (2023), Privacy- and Utility-Preserving NLP with Anonymized data: A case study of Pseudonymization, in *Proceedings of the 3rd Workshop on Trustworthy Natural Language Processing (TrustNLP 2023)*, Toronto, Association for Computational Linguistics, pp.232-241

Patti L. (2019), Art. 58/Regolamento UE 2016/679, in D'Agostino A., Barlassina L.R., Colarocco V. (a cura di), *Commentario al Regolamento UE 2016/679 e al Codice della Privacy aggiornato*, Vignate, Top Logal Academy, pp.460-468

Ponti B. (2023), *Attività amministrativa e trattamento dei dati personali,* Milano, Franco Angeli

Simonati A. (2023), Il trattamento di dati personali e gli accessi amministrativi "generali": le dinamiche frontiere della discrezionalità, *Diritto Amministrativo*, 31, n.1, pp.3-27

Stalla-Bourdillon S. (2025), The state of pseudonymisation in the EU: Where do we stand today?, *Privacy and Data Protection,* 18, n.2, pp.4-8

Stefanelli S. (2018), Trattamento di dati personali per scopi di ricerca scientifica, in Cassano G., Colarocco V., Gallus G.B., Micozzi F.P. (a cura di), *Il processo di adeguamento al GDPR*, Milano, Giuffré Francis Lefebvre, pp.369-399

Trubiani F. (a cura di) (2025), *Sistemi di intelligenza artificiale in medicina verso lo spazio europeo dei dati sanitari*, Giappichelli, Torino

Chiara di Somma

C.DiSomma@gpdp.it

Giurista esperta in protezione dei dati personali, funzionario presso l'Autorità Garante per la protezione dei dati personali dal 2009. Nel corso della sua formazione ha conseguito un Master di II Livello in Tutela dei Diritti umani presso l'Università Sapienza di Roma (2004) ed un Dottorato di ricerca in Diritto pubblico e costituzionale presso l'Università Federico II di Napoli (2008). Oltre alle numerose docenze nell'ambito di Master universitari e corsi di formazione in tema di protezione dei dati personali è cultrice della materia presso l'Università Luiss Guido Carli (corso di Diritto digitale e tutela dei dati), Corso di Laurea Magistrale in Giurisprudenza A.A. 2025-2026. È inoltre autrice di numerose pubblicazioni scientifiche in ambito giuridico nel settore della protezione dei dati personali.

Note su avventure e sventure dei dati amministrativi, con riferimento all'esperienza di Veneto Lavoro

Bruno Anastasia
Esperto di analisi del mercato del lavoro

Il saggio analizza opportunità e ostacoli nell'uso dei dati amministrativi per la conoscenza del mercato del lavoro, ricostruendo la concreta esperienza trentennale dell'Osservatorio di Veneto Lavoro: dalla raccolta cartacea alla digitalizzazione con Netlabor, fino alla svolta delle Comunicazioni Obbligatorie telematiche (2008). Affronta nodi metodologici (relazione flussi-stock) e organizzativi (rapporti interistituzionali, apertura a ricercatori esterni tramite Public Use File), fino a considerare le più recenti potenzialità connesse al ruolo di SIU e SIISL e all'analisi degli utenti dei Centri per l'impiego.

Notes on the uses and challenges of administrative data: the experience of Veneto Lavoro

This essay examines the opportunities and obstacles in using administrative data to understand the labour market, reconstructing the thirty-year experience of the Veneto Lavoro Observatory: from paper-based data collection to digitalisation through Netlabor, up to he turning point represented by mandatory electronic Employment Communications (2008). It addresses methodological issues (the flow-stock relationship) and organisational challenges (inter-institutional relations and access for external researchers through Public Use Files), and discusses the recent potential linked to the role of SIU and SIISL and to the analysis of Public Employment Service users.

DOI:10.53223/Sinappsi_2026-02-2

Citazione
Anastasia B. (2026), Note su avventure e sventure dei dati amministrativi, con riferimento all'esperienza di Veneto Lavoro, *Sinappsi*, XVI, n.2, pp.34-48

Parole Chiave
Dati amministrativi
Integrazione dei dati
Mercato del lavoro

Keywords
Administrative data
Data integration
Labour market

Introduzione

Le possibilità di utilizzo dei dati amministrativi a fini conoscitivi in materia sociale ed economica hanno aperto opportunità crescenti, che però risultano strettamente collegate e condizionate – oltre che da fattori tecnologici (digitalizzazione ecc.) – dall'evoluzione istituzionale e dal comportamento degli attori pubblici, con riferimento in particolare alla loro interpretazione della 'mission' che devono assolvere, con i conseguenti risvolti in materia di concezione patrimoniale dei dati posseduti (e relative gelosie/chiusure) e di limiti veri o presunti dettati dalla normativa sulla privacy. A tal fine, dopo qualche considerazione sui rischi di overdose di dati statistici (paragrafo 1) e sulla rilevanza e specificità dei dati amministrativi (paragrafo 2), ci si sofferma (paragrafo 3) sul *case-study* costituito dall'esperienza di Veneto Lavoro – lungo circa un trentennio – in materia di trattamento dei dati amministrativi relativi al mercato del lavoro (comunicazioni obbligatorie delle imprese a proposito di assunzioni, trasformazioni, cessazioni; dichiarazioni di disponibilità da parte dei disoccupati ecc.). Si propone infine una considerazione conclusiva su quale sarebbe il buon uso auspicabile dei dati nella discussione pubblica.

1. Siamo ancora sicuri che valga la pena aumentare la produzione statistica?

È usuale enfatizzare l'importanza dei 'dati' (numeri associati a determinati fenomeni) nella società attuale. Ciò vale anche per quel gruppo specifico di 'dati' che costituisce la materia prima per l'elaborazione di statistiche e connesse valutazioni/interpretazioni su società ed economia. Si dice 'dati': si intende statistiche, misure, conteggi. Di fatto svolgono due diverse tipologie di servizio:

a. *a fini conoscitivi: contribuiscono in modo importante a formare – tanto negli utenti-destinatari quanto negli stessi produttori – l'idea di come va il mondo.* Siamo naturalmente curiosi e vogliamo sapere se la popolazione della Francia è più numerosa di quella della Germania, se lo stipendio medio degli insegnanti in Spagna è inferiore a quello degli insegnanti in Olanda ecc. Viviamo in un mondo (per ora) iperconnesso, globalizzato, sicuramente grande: per farci qualche idea delle dimensioni relative di redditi, popoli, strutture ecc. dobbiamo aver a disposizione dati statistici, fotografie fredde, impersonali ma efficaci. Che non servono – o servono a poco – per farci un'idea delle realtà che si misurano con gli occhi e con l'esperienza diretta (il benessere di un piccolo paese di campagna, ad esempio) ma per dire qualcosa dell'Italia o, ancora di più, della Cina, gli occhi non bastano e le statistiche diventano irrinunciabili. Sulla loro base ci si fa e si comunica l'idea di come va il mondo, di quali siano i pericoli (l'inverno demografico), le variazioni (se la povertà cresce, se la disuguaglianza tracima), le opportunità (le chance di impiego o di business). Finché non viene statisticato (più o meno bene), un fenomeno sociale ottiene poca audience. Questo ruolo importantissimo delle statistiche nel formare/condizionare le opinioni su come va il mondo e su come dovrebbe/potrebbe andare fa sì che esse non siano mai materiale neutro: nessuna meraviglia, dunque, se vengono usate e abusate nel discorso pubblico, con ricadute dirette nell'arena politica. Data questa finalità essenzialmente di carattere conoscitivo generale ci può essere ampia tolleranza per numeri di bassa qualità, imprecisi e parziali: essi servono, infatti, essenzialmente alle discussioni, ai confronti, ai convegni, alla propaganda (più o meno onesta e seria), con ricadute indirette sull'operatività;
b. *a fini operativi: accade quando le statistiche vengono assunte come base per decisioni e azioni concrete degli operatori pubblici e privati*. Ad esempio, se il determinato tasso di disoccupazione di un territorio consente l'accesso delle sue imprese a determinate agevolazioni fiscali. In tal caso il dato statistico non serve solo alla conoscenza ma anche – o soprattutto – ad indirizzare flussi di risorse con un impatto materiale importante.

Con riferimento alla finalità conoscitiva, le distorsioni nella comprensione e nell'utilizzo delle statistiche non sono rare. La radice di queste distorsioni sta non solo nelle semplificazioni[1] finalizzate in genere ad un uso selettivo dei numeri in funzione delle proprie, già formate, convinzioni ma anche nell'éscamotage di addossare sull'interpretazione delle statistiche il peso di divergenze che in realtà vertono su obiettivi di policy che non si riesce (o, più raramente, non si vuole) focalizzare ed esplicitare[2].

Se così stanno le cose, ha senso auspicare e aumentare la produzione di statistiche (più disaggregate, più granulari, più tempestive), visto che le fonti amministrative, a causa della digitalizzazione delle loro procedure, lo consentono a basso costo?

La risposta non è scontata[3]. Rispondere positivamente implica un atto di fiducia nella democrazia, quantomeno augurarsi – se non proprio essere convinti – che nello spazio pubblico la formazione delle opinioni e delle decisioni possa avvenire nel modo più serio, documentato e il più trasparente possibile. Per tale processo le statistiche sono un carburante necessario. Per questo sono un bene pubblico e la loro ricchezza e qualità è patrimonio collettivo importante.

1 Spesso alimentate dall'idea balzana che 'i numeri parlano da soli': in realtà ciò non avviene quasi mai, serve sempre chiarirne perimetro, attendibilità, definizioni utilizzate. Luciano Capone ha opportunamente osservato, partendo da una recente querelle intorno alla parola 'cortigiana', che "nel dibattito pubblico puoi anche diffondere in continuazione dati sballati ma non puoi permetterti di usare termini ambigui. Insomma, devi stare ben attento alle parole, ma sui numeri puoi fare come ti pare. (…) Se è molto probabile trovare un interlocutore che immediatamente ti corregge dicendo 'occhio, che questo termine è un po' sessista', è praticamente impossibile beccarne uno che (...) ti dica 'attenzione, che queste cifre non sono vere'" (*Il foglio,* 21 ottobre 2025).

2 Occorre tener conto del paradosso per cui l'arte della politica è la capacità di prendere decisioni (giuste) in assenza (o in carenza) di informazioni. Se circa un determinato fenomeno sociale e la sua traiettoria evolutiva fossero disponibili tutte le informazioni utili al suo governo, qualsiasi attore prenderebbe la stessa decisione, a meno che non vengano esplicitati obiettivi diversi: solo essi giustificherebbero decisioni diverse.

3 Si veda Nespoli (2025). Sul destino dei dati di essere "maltrattati" cfr. Anastasia (1993).

2. La rivoluzione dei dati amministrativi

Le trasformazioni tecnologiche che si sono imposte su larga scala nell'ultimo ventennio del secolo scorso – informatizzazione (archivi elettronici), digitalizzazione (formati strutturati) e Internet (diffusione via web) – hanno profondamente modificato produzione e distribuzione delle statistiche, anche con riferimento al lavoro, all'occupazione e alla disoccupazione.

Sostanzialmente fino agli anni Ottanta i dati disponibili sul mercato del lavoro italiano sono stati, pressoché esclusivamente, quelli derivanti dall'Indagine Istat sulle Forze di lavoro: indagine campionaria avviata fin dagli anni Cinquanta, finalizzata alla produzione quanto più tempestiva dei dati utili alle politiche keynesiane (tasso di disoccupazione in primis). Con le ovvie limitazioni implicite in un'indagine campionaria, per quanto accurata e basata su un campione molto largo: l'impossibilità di disaggregazioni territoriali, settoriali, professionali 'fini'. Contemporaneamente il mondo del lavoro era interessato da notevole attività amministrativa: gli Uffici di collocamento accertavano la disoccupazione e gli avviamenti al lavoro[4]; l'Inps registrava i versamenti contributivi delle imprese a favore dei dipendenti nonché i versamenti dei lavoratori autonomi. Le potenzialità statistiche di queste attività amministrative erano però largamente trascurate, perché troppo onerose da concretizzare (implicavano attività manuali di conteggio ecc.) e ritenute difficili da veicolare per obiettivi di conoscenza[5]. E in effetti è pressoché impossibile trovare, almeno fino agli anni Novanta, importanti ricerche e analisi sul mercato del lavoro basate su dati amministrativi[6].

È solo verso la fine del secolo che si avvia lo sfruttamento dei dati amministrativi[7] e si diffonde la consapevolezza del loro vantaggio competitivo, dovuto a diversi fattori:

a. i bassi costi. I dati amministrativi non implicano la realizzazione di ricerche ad hoc (come le indagini campionarie) con tutte le connesse problematiche di costo, se si vuol fare un buon campionamento e una rilevazione accurata, perché vengono prodotti come risultato collaterale di un'attività finalizzata ad altro e perciò già finanziata. Sono come la segatura rispetto al processo di produzione di mobili: essa non è la finalità del processo produttivo, cionondimeno viene prodotta e può essere riciclata e venduta, con tutti i conseguenti vantaggi;
b. la possibilità di disaggregazioni coerenti e utilizzabili, dato che i dati amministrativi sono per definizione universali (relativamente alla platea prevista dalle norme) e non soffrono quindi i limiti al riguardo delle indagini campionarie[8];
c. l'utilizzabilità diretta a fini di policy. Da un'indagine campionaria si ricava il numero dei disoccupati (o dei poveri ecc.) ma non ne segue la possibilità di identificarli singolarmente: si stimano dei numeri senza conoscere i nomi; in sede di policy si possono predisporre programmi di intervento ma non si dispone di un'identificazione precisa della platea di beneficiari. Con i dati amministrativi tutto ciò non accade;
d. il più frequente riferimento a classificazioni comunemente note, che non di rado differiscono da quelle della statistica ufficiale. Spesso ad esempio i non addetti ai lavori ritengono che i numeri dei disoccupati ufficiali (quelli riportati da tv e media) sono riferiti ai classici 'iscritti al collocamento' (oggi divenuti i 'disponibili secondo le liste dei Centri per l'impiego'): è sempre stato difficile spiegare che, invece, derivano da un'indagine campionaria riconosciuta come statistica ufficiale; lo stesso vale per diversi aggregati e indicatori della contabilità nazionale (pressione fiscale, economia sommersa ecc.) la cui trasposizione rispetto a ciò che si osserva e si conosce dalla vita quotidiana non è immediata né semplice;
e. la valorizzazione dei dati di flusso, oltre – spesso prima di – quelli di stock. Derivando

4 La regolazione amministrativa del mercato del lavoro prevedeva che per essere assunti si dovesse essere necessariamente disoccupati (almeno per un giorno).

5 Per una trattazione ampia e un inquadramento storico preciso della storia delle statistiche sul lavoro, con particolare attenzione a quanto impostato negli anni Cinquanta, si veda Favero e Trivellato (2000). Un aggiornamento fino alle vicende del primo decennio del secolo è proposto in Sestito e Trivellato (2010).

6 Si segnalano comunque i sei corposi rapporti redatti tra il 1987 e il 1994 dalla Fondazione Brodolini per conto del Ministero del Lavoro e della previdenza sociale, in adempimento di quanto previsto dalla legge n. 56/87, intitolati *Lavoro e politiche dell'occupazione in Italia*: i dati amministrativi vengono utilizzati per documentare specifici provvedimenti (Cassa integrazione, apprendistato, qualche politica del lavoro) ma non per ricavarne indicazioni generali (men che mai sulla disoccupazione) né integrazioni/interlocuzioni con i dati provenienti dalle fonti campionarie ufficiali.

7 Si intende sempre, in questa nota, dati amministrativi di istituzioni pubbliche.

8 Talvolta si utilizzano anche campioni (grandi campioni) di dati amministrativi, soprattutto per economizzare il necessario lavoro di elaborazione. In tal modo si ritorna però ai limiti tipici delle indagini campionarie, anche se attenuati da strategie di campionamento ampiamente rappresentativo.

da attività istituzionali continue nel tempo, i dati amministrativi generano per definizione innanzitutto dati di flusso. Spesso per la programmazione e la realizzazione delle policy, i dati di flusso sono più rilevanti di quelli di stock (es. quanti dipendenti vengono licenziati in un anno, quanti rilasciano la dichiarazione di disponibilità ai Centri per l'impiego, quanti accedono all'assicurazione contro la disoccupazione (NASpI ecc.). Le indagini campionarie sono adatte invece soprattutto a scattare fotografie, istantanee[9]. Solo disponendo dei dati di flusso possiamo conoscere le dimensioni di tutti i cambiamenti intercorsi tra un Censimento e l'altro: è come disporre di un filmato che mostra tutti i cambiamenti intervenuti in un dato periodo, cambiamenti che sono la spiegazione delle differenze tra la fotografia iniziale e quella finale;

f. la possibilità di agevoli elaborazioni longitudinali, importanti per distinguere l'esperienza di una data generazione nel tempo dal confronto sezionale tra due momenti storici distinti. Infatti, l'analisi longitudinale può dare risultati che differiscono significativamente da quelli cross-section, derivanti dal confronto tra platee differenti nel tempo per composizione ed estensione. Ad esempio, nel caso di variazioni retributive, per la generazione X entrata nell'anno t si potrebbe evidenziare nell'anno t+1 una reale crescita salariale (nel valore medio del salario) e ciò potrebbe benissimo non coincidere con il risultato del confronto tra le retribuzioni di tutti i salariati in t+1 rispetto a tutti i salariati in t[10].

A fronte di questi innegabili vantaggi, vanno ricordati anche gli ostacoli specifici all'utilizzo su larga scala dei dati amministrativi. Tali ostacoli dipendono soprattutto dalla loro natura originaria, vale a dire dal fatto di essere stretta proprietà di specifiche istituzioni pubbliche (Ministero o Regione o Ente previdenziale) che non hanno, statutariamente, la *mission* di fornire informazioni statistiche all'opinione pubblica a scopi generali di conoscenza[11]. Se lo fanno, è per loro iniziativa (bontà) ed entro i limiti che si autoassegnano per ragioni di varia natura, a volte giustificate – come l'attenzione a non interferire con ambiti di intervento di altre amministrazioni, specie se sovraordinate – altre volte meno: i dati sono anche una cartina di tornasole, che può risultare antipatica, della qualità dell'attività amministrativa, dell'accuratezza delle metodologie di lavoro[12], dell'avanzamento a proposito di digitalizzazione e innovazione. Le informazioni statistiche non sono così innocue come sembra: quando non accurate, lasciano intravedere in controluce le debolezze degli enti proprietari, che si preferisce restino nascoste, non certo esposte in pubblico[13].

Last but non least, la questione della privacy e i timori per l'invadenza del Garante: la privacy è di fatto il miglior assist per quelle Amministrazioni pubbliche desiderose di centellinare le informazioni da mettere a disposizione di tutti.

3. L'esperienza dell'Osservatorio di Veneto Lavoro

L'esperienza dell'Osservatorio di Veneto Lavoro – in continuità con l'attività già avviata presso l'Agenzia per l'impiego del Veneto[14] – è stata tutta inscritta dentro la convinzione di poter raggiungere buoni risultati nell'analisi del mercato del lavoro e delle politiche pubbliche ad esso riferite lavorando soprattutto sui bistrattati dati amministrativi, di proprietà allora – nei primi anni Novanta – del Ministero del Lavoro e delle sue articolazioni territoriali – Uffici periferici del lavoro, Sezioni circoscrizionali per l'impiego, il collocamento

9 Erano un'istantanea anche i dati dei Censimenti decennali svolti fino al 2011: fotografie di specifici momenti senza la possibilità di conoscere le dimensioni dei movimenti intervenuti tra un Censimento e l'altro, che rimanevano oggetto di difficili stime.

10 Ciò a causa di diversi assetti nella composizione delle platee: per esempio potrebbe dipendere dal fatto che gli entrati in t+1 hanno beneficiato mediamente di una retribuzione inferiore a quella degli usciti (presenti solo in t).

11 Qualche eccezione (forse) va fatta per le statistiche previste dal triennale Programma statistico nazionale (attualmente è in corso di definizione il Programma 2026-2028).

12 Come è ben noto: *garbage in, garbage out*.

13 È il caso, ad esempio, del dibattito di questi mesi/anni sulle liste di attesa in sanità. Si tratta di semplici informazioni statistiche, di origine amministrativa, di cui si teme molto la ricaduta negativa sul prestigio dei dirigenti delle aziende sanitarie, delle competenti strutture regionali e alla fine del Ministero della Sanità. Per questo si ingigantiscono oltremodo i problemi della raccolta/trattamento/diffusione dei dati relativi (e, se proprio si deve produrre queste informazioni, si punta quantomeno a limitarne la diffusione e l'accessibilità).

14 Nate nei primi anni Novanta (quella del Veneto nel 1992), in anticipo rispetto alla riforma Bassanini (1997), le Agenzie regionali per l'impiego dovevano rispondere ad esigenze di sperimentazione organizzativa e di implementazione di capacità di analisi e programmazione. Di fatto hanno anticipato il lessico e alcune funzioni del sistema decentrato pur restando diretta emanazione delle strutture centrali dello Stato, in sostanza un ibrido: statali nella titolarità, regionali nella scala operativa.

e l'assistenza (SCICA) – oggi direttamente del sistema regionale: Regione, Veneto Lavoro, Centri per l'impiego (CPI), come esito di successive trasformazioni istituzionali[15].

A cavallo tra gli anni Ottanta e gli anni Novanta del secolo scorso, importanti lavori di un gruppo di ricercatori torinesi – in primis Riccardo Revelli e Bruno Contini – avevano imposto all'attenzione di tutti i ricercatori interessati alle questioni del mercato del lavoro la ricchezza dei dati amministrativi Inps[16]. Perché non provare a seguirne le orme con i dati dei Centri per l'impiego? È così che si è cominciato a lavorare, cercando sempre riscontri e confronti con le fonti principali (Istat in primis, ma anche Inps man mano che risultava possibile). Con la convinzione che la realtà del mercato del lavoro è una sola, seppur indagabile da diverse angolature e con fasci di luce rivolti in diverse direzioni. La legittimità dell'uso dei bistrattati dati dei CPI poteva solo essere rafforzata dall'interazione e dal confronto con le altre informazioni disponibili sul mercato del lavoro.

I CPI possedevano e possiedono due grandi gruppi di informazioni: a. *sui rapporti di lavoro*: le Comunicazioni Obbligatorie (d'ora in poi CO) dei datori di lavoro per tutti i movimenti attivati (assunzioni, cessazioni, trasformazioni); b. *sui disoccupati*: le dichiarazioni di disponibilità e la conseguente attività orientata alle politiche attive (tirocini, formazione, accompagnamento ecc.).

La possibilità di lavorare su questi importanti temi di ricerca è mutata nel tempo, via via ampliandosi, in parallelo all'evoluzione delle tecnologie, dell'assetto istituzionale e della conseguente domanda di conoscenza.

Di seguito ricapitoliamo alcuni passaggi particolarmente importanti.

La situazione iniziale: quando il supporto era solo la carta

Agli esordi, l'Agenzia del Veneto raccoglieva dalle SCICA i tradizionali modelli ministeriali (Oml1, Oml2 ecc.) compilati su supporto cartaceo: essi venivano digitalizzati a cura dell'Agenzia stessa per essere commentati soprattutto nel *Rapporto annuale*[17]. Solo di qualche specifico programma di intervento (come le *Liste di mobilità* previste dalla L. n. 223/1991 che avevano innovato la regolazione dei licenziamenti collettivi) l'Agenzia era titolare anche dei microdati, con conseguenti più ampie possibilità di elaborazioni. Ma già il fatto che le statistiche tradizionali, frutto di computo manuale degli operatori delle SCICA, venissero raccolte, aggregate su scala regionale, controllate e divulgate, induceva i produttori a maggiore attenzione e ad aumentare quindi l'affidabilità del risultato, tradizionalmente ritenuto carente data la certezza da parte degli operatori-produttori che nessuno avrebbe degnato quei dati di soverchia attenzione. I primi test di raffronto con i dati Istat sulle dinamiche del lavoro dipendente davano risultati confortanti, oltre le aspettative, attestando che – se trattata adeguatamente – la fonte era di indubbio interesse[18].

La diffusione di Netlabor, la prima digitalizzazione

L'informatizzazione dei CPI, sviluppatasi in Veneto nel corso degli anni Novanta, a partire soprattutto dal basso, con la diffusione del software Netlabor[19] – essendosi arenati i progetti centrali del Ministero – si è completata attorno al 1998 con un forte impatto sul miglioramento dell'attività amministrativa (adempimenti e archiviazione). L'Osservatorio colse fin dall'inizio l'opportunità che si apriva di

15 Il primo passaggio decisivo è stato nel 1997 il decreto 469 (riforma Bassanini): prevedeva il superamento delle SCICA (Sezioni circoscrizionali per l'impiego, il collocamento e l'assistenza: avevano preso il posto dei vecchi Uffici comunali di collocamento con la L. n. 56/1987) e l'avvento dei Centri per l'impiego (CPI), la cui gestione operativa veniva affidata alle Province mentre le Regioni divenivano titolari delle funzioni in materia di collocamento e politiche attive, assorbendo anche le Agenzie regionali per l'impiego. Nel 1998 la Regione Veneto ha istituito l'ente Veneto Lavoro (L. n. 31/1998) attribuendo al nuovo ente tutte le funzioni di osservazione e monitoraggio del mercato del lavoro. Un secondo passaggio cruciale è stato avviato nel 2014 (riforma Del Rio) e perfezionato nel 2015 con il *Jobs Act* (D.lgs. n. 150/2015): i Servizi per l'impiego sono divenuti funzione propria delle Regioni e quindi i CPI sono stati assorbiti dalle Regioni direttamente o tramite enti strumentali regionali (come, nel caso del Veneto, da Veneto Lavoro).

16 Tappa fondamentale è stato il volume di Contini e Revelli (1992), *Imprese, occupazione e retribuzioni al microscopio: studi sull'economia italiana alla luce delle fonti statistiche INPS.*

17 La prima edizione del *Rapporto annuale* è stata presentata nel 1993; è stato pubblicato regolarmente sempre da Franco Angeli fino al 2013. L'ultima edizione, disponibile nel sito di Veneto Lavoro (https://www.venetolavoro.it/documents/10180/2353634/Rapporto_2014.pdf) è quella del 2014 (*Discesa finita? Crisi al sesto anno: assottigliati ancora posti di lavoro e risorse imprenditoriali*).

18 Come dimostrato in Anastasia *et al.* (2009). Questo lavoro presenta un'ampia analisi dell'evoluzione della fonte con il passaggio dal cartaceo alla digitalizzazione, passaggio legato ai processi di informatizzazione dei CPI.

19 Progettato e saggiato in Veneto, a cura di un Gruppo informatico costituito fin dai primi anni Novanta e composto da un referente per ogni provincia, Netlabor è stato progressivamente distribuito, grazie anche a un impegno in tal senso del Ministero del Lavoro, a gran parte degli uffici periferici dell'intero Paese. Sulle vicende legate alla prima informatizzazione degli uffici e all'analisi delle ricadute per l'attività di ricerca e monitoraggio cfr. Anastasia e Trivellato (2001) e Bassi *et al.* (2021).

disporre di microdati per ricavarne informazioni svincolate dalle gabbie dei tradizionali modelli ministeriali di raccolta dei dati (che sopravvivevano in parallelo alla prima diffusione di Netlabor). Grazie alla collaborazione con la Direzione provinciale di Venezia, si è testato l'utilizzo di microdati per esplorare i flussi (assunzioni, trasformazioni, cessazioni) in un sistema locale del lavoro caratterizzato da forte stagionalità (Anastasia *et al.* 1996). Data la verifica positiva, l'esperienza maturata è stata affinata e rivolta sia a indagare specifici temi di studio sia a identificare una possibile produzione standardizzata di dati utili all'analisi complessiva del mercato del lavoro. Si è dato vita quindi a progetti di costruzione di nuove statistiche di base valorizzando il patrimonio informativo di Netlabor. In tal senso andavano il Progetto Val.net e i collegati Progetto Quick (per l'analisi congiunturale) e Progetto A.Si. Lo. (Analisi Sistemi Locali): da essi deriva la proposta di strutturazione del monitoraggio dei mercati del lavoro presentata alla Quarta Conferenza nazionale di statistica del 1998, congiuntamente con l'Ufficio del Dirigente responsabile dei Sistemi informativi automatizzati del Ministero del Lavoro.

Negli anni successivi, ci si è dovuti confrontare con i numerosi e radicali problemi derivanti dagli assetti istituzionali in transizione e soprattutto da un'informatizzazione che procedeva a pelle di leopardo e ripetutamente destabilizzata dall'esigenza di internalizzare le possibilità offerte dall'evoluzione tecnologica (passaggio a Netlabor4 con sistema di replica parziale dei dati dei singoli CPI su un server centrale). Di fatto ogni CPI faceva storia a sé, con livelli diversi di aggiornamento, completezza e accuratezza delle informazioni digitalizzate: in genere migliori per i CPI di piccole dimensioni, sempre parziali e/o in ritardo per i CPI delle grandi città. Ciò rendeva impossibile la ricostruzione tempestiva di basi dati stabili a livello regionale: anche l'individuazione di un gruppo di CPI fornitori di dati di qualità si rivelò improduttiva perché invalidata da troppe entrate-uscite[20].

Tutto ciò ha determinato la chiusura del Progetto Val.net, orientato al monitoraggio congiunturale, e a sviluppare, invece, attività orientate a utilizzi delle basi dati a fini di ricerca, sfruttando anche le potenzialità della fonte per lo sviluppo di analisi longitudinali. In questo ambito si sono ottenuti buoni risultati indagando specifici temi, compatibili con il livello esistente di digitalizzazione. Di questa attività, oltre che nelle varie pubblicazioni di Veneto Lavoro di quegli anni, importanti riscontri sono rintracciabili nel consistente contributo dato dall'Osservatorio Veneto al *Gruppo di lavoro per il monitoraggio degli interventi di politica occupazionale e del lavoro*, istituito con Decreto del Ministro del Lavoro e della previdenza sociale il 17 dicembre 1999, guidato da Paolo Sestito (cfr. Ministero del Lavoro e della previdenza sociale 2006), come pure dalla partecipazione a importanti progetti di ricerca in ambito universitario[21].

La svolta: l'avvento delle CO telematiche

Un vero spartiacque è stato il biennio 2007-2008, con un insieme di riforme poco raccontate, apparentemente burocratiche e marginali, che in realtà hanno posto le premesse per maggior trasparenza, più controllo, più conoscenza del mercato del lavoro, con implicazioni decisive per le attività sia di osservazione che (almeno potenzialmente) di intervento sul mercato del lavoro: si tratta del passaggio all'obbligatorietà delle comunicazioni delle imprese in forma telematica nonché dell'estensione della platea dei datori di lavoro tenuti a tali comunicazioni[22]. Inoltre, l'obbligo di comunicazione è stato esteso anche ai rapporti di lavoro di tipo parasubordinato (collaborazioni, associazioni in partecipazione) come pure alle esperienze di lavoro (tirocini non curriculari, lavori socialmente utili)[23]. Fino ad allora il sistema era frammentato (comunicazioni diverse a Inps, Inail, Centri per l'impiego), antiquato (cartaceo o semi-cartaceo), facilmente eludibile soprattutto nel caso di rapporti brevi.

Il passaggio chiave è stata l'attuazione della riforma introdotta con il decreto legislativo n.

20 I problemi di quegli anni sono accuratamente descritti nel già citato Anastasia *et al.* (2009).

21 Si veda, ad esempio, il contributo dato a programmi di ricerca cofinanziati dal MIUR: Contini e Trivellato (2005); va segnalata anche l'ampia ricerca sulla diffusione del tempo determinato svolta in collaborazione con Aris Accornero (coordinatore) ed Elisabetta Gualmini (Agenzia per l'Impiego del Veneto 2000).

22 Dal 1 gennaio 2007 l'obbligo di comunicazione in materia di rapporti di lavoro dipendente è stato esteso anche alle Pubbliche amministrazioni.

23 Da queste riforme era rimasto escluso il lavoro domestico, per il quale comunque rimaneva l'obbligo di comunicazione ai Centri per l'impiego (modello Unilav/lavoro domestico). Successivamente è stato stabilito che dal 29 gennaio 2009 tali comunicazioni dovevano essere fatte all'Inps; poi dal 1 aprile 2011 è stato introdotto l'obbligo di comunicazione, sempre in via telematica, comunque all'Inps (che poi le inoltra ai CPI). È un esempio dell'irresistibile tendenza alla frammentazione e sovrapposizione di compiti, per cui ogni disegno di riforma non è mai limpido e completo. Si possono citare anche altri casi: vedi ad esempio l'obbligo di dimissioni online introdotto nel 2016 con ovvia costruzione di altra distinta piattaforma; oppure le incertezze sull'obbligo di comunicazione nei casi di passaggio dal full al part-time (da rilasciare all'Ispettorato del lavoro anziché ai Centri per l'impiego), su cui si veda Anastasia e Maschio (2017).

181/2000, poi modificato dal D.Lgs. n. 297/2002, resa finalmente pienamente operativa con il DM 30 ottobre 2007: dal 1 marzo 2008 tutti i datori di lavoro sono obbligati ad un'unica comunicazione alle diverse strutture del sistema pubblico interessate (Ministero del Lavoro, Inps, Inail)[24]. Per gran parte delle assunzioni – fanno eccezione i rapporti di somministrazione e il settore pubblico[25] – si tratta di una comunicazione ex ante: potenzialmente rende tracciabile in tempo reale l'andamento delle assunzioni e con un brevissimo lag temporale – dovuto ai tempi concessi per la comunicazione delle cessazioni – anche la dinamica dello stock di posizioni di lavoro. Inoltre, l'obbligo di comunicazione telematica preventiva alza significativamente rischio – e quindi costo – della trasgressione.

La svolta del 2008 avviene, si può dire fortunatamente ai fini del monitoraggio, in contemporanea all'esplodere della recessione internazionale originata in USA dai sub-prime, diffusasi su scala globale soprattutto dopo il fallimento Lehman Brothers (15 settembre 2008), con una particolare successiva accentuazione in Italia a seguito dei rischi derivanti dalla crisi dei debiti sovrani dei Paesi mediterranei (2011-2012).

L'obbligo di comunicazione telematica da parte dei datori di lavoro azzera le differenze tra un CPI e l'altro, superando le lentezze connesse alla digitalizzazione delle informazioni da parte degli operatori dei Centri. Diventa immediatamente possibile il monitoraggio tempestivo degli andamenti sul mercato del lavoro: per questo la crisi occupazionale in Veneto è stata monitorata tempestivamente e dettagliatamente, per area e per settore. Molto più difficile è stato monitorare adeguatamente e tempestivamente gli strumenti messi in campo per arginarla[26].

Con l'avvento delle CO telematiche sia lo Stato che ciascuna Regione sono in grado di effettuare analisi e monitoraggi assai dettagliati e approfonditi. Concretamente le esperienze sono state (e sono) le più disparate: si va dal quasi nulla (per varie ragioni: disinteresse dei dirigenti, incomprensione della potenzialità, mancati investimenti in competenze e risorse umane adeguate) a un uso molto intenso, come nel caso del Veneto[27], con una vasta produzione messa a disposizione sia degli attori più direttamente interessati (operatori politici e sindacali, amministratori) sia dell'opinione pubblica.

Nonostante le riconosciute potenzialità, a livello nazionale fino al 2015 le CO non sono state praticamente utilizzate a fini di Osservatorio. Questo vuoto è stato parzialmente coperto – e, implicitamente, sottolineato – da un'iniziativa autonoma di coordinamento dal basso che ha portato alla costituzione del network Statistiche e Comunicazioni Obbligatorie (Se.C.O.). Avviato nel 2011 per iniziativa dei ricercatori di alcune regioni italiane, essenzialmente del Centro-Nord (Piemonte, Lombardia, Liguria, Veneto, Friuli Venezia Giulia, Emilia-Romagna, Marche, Umbria, Campania, Sardegna e le Province autonome di Trento e Bolzano), il network si era dato l'obiettivo di armonizzare e condividere metodologie comuni di trattamento e restituzione dei dati ricavabili dalle CO. Sono stati continuativamente prodotti, fino al 2015, report trimestrali messi online[28] dai singoli Osservatori regionali o provinciali. Si è trattato di un'attività lontana dagli esigenti standard richiesti per una produzione statistica di assoluta qualità[29] ma comunque rilevante: sia perché segnalava implicitamente che a livello centrale si stava perdendo un'opportunità, sia perché visibilizzava l'impegno dei partecipanti a condividere metadati e a confrontare i risultati ottenuti.

L'attenzione alle questioni metodologiche: il legame cruciale tra flussi e stock

La disponibilità dei dati di CO ha imposto il problema della relazione tra flussi e stock. Le CO raccolgono informazioni su flussi: assunzioni, cessazioni e trasformazioni di ogni tipo dei rapporti di lavoro. Ma tutte le conoscenze standard sul mercato del lavoro riguardano stock: occupati, unità di lavoro

24 La storia tecnologica e istituzionale delle CO meriterebbe di essere accuratamente ricostruita: tra il dire (comunicazione unica a tutti gli Enti) e il fare (scambio di informazioni tra tutti gli Enti, interoperabilità, utilizzabilità sia a fini operativi che di conoscenza) c'è di mezzo il mare dei conflitti e delle gelosie istituzionali (in particolare: i dati sono ugualmente di tutti? o qualcuno ha la possibilità di chiudere/aprire il rubinetto? ecc.) e delle tensioni sempre possibili tra chi presiede alle scelte informatiche e chi ha il compito di utilizzare i dati per esigenze operative o conoscitive. Comunque, nonostante ritardi e problemi, le CO sono diventate uno dei primi grandi sistemi nazionali digitali realmente funzionanti.

25 Al settore pubblico è concessa, senza motivazioni convincenti, la possibilità di effettuare le comunicazioni entro il ventesimo giorno del mese successivo a quello dell'evento; le agenzie di somministrazione sono tenute alla comunicazione entro i 5 giorni successivi all'evento.

26 All'impatto decisivo della Cig in deroga è stata comunque dedicata una specifica, quasi pionieristica, ricerca: Anastasia e Basso (2015).

27 E non solo: vi è un'attività evoluta in diverse regioni del Centro-Nord mentre meno brillante risulta la situazione al Sud.

28 Tuttora disponibili ad esempio in https://www.venetolavoro.it/seco1.

29 Come descritti ad esempio in Istat (2025).

e loro variazioni. Non si tratta però di prospettive – e di statistiche – inconciliabili. Anzi, è del tutto logico attendersi che, controllati i possibili fattori di differenziazione (perimetro, definizioni, unità di analisi), la conciliazione tra stock e flussi emerga da sé: si tratta infatti di misure necessariamente integrate e concordanti. Se non avviene è necessario spiegarne le ragioni. Troppo comodo tenerle separate, sottraendole così alla valutazione di qualità ed esimendo i produttori dalla ricerca di eventuali punti deboli. L'Osservatorio di Veneto Lavoro ha sempre prestato estrema attenzione sia all'analisi distinta dei flussi sia alla possibilità di calcolare, trattando opportunamente i dati, non gli stock (per definizione ignoti, dato l'avvio ancora relativamente recente del sistema CO[30]), quanto le variazioni degli stock. Variazioni che possono essere calcolate anche per specifici sub-aggregati (ad esempio le tipologie contrattuali), integrando opportunamente tutte le informazioni disponibili: ad esempio, per ricostruire correttamente la variazione dello stock di posizioni a tempo indeterminato, occorre tener conto, oltre che di assunzioni e cessazioni, anche delle trasformazioni da tempo determinato a indeterminato.

Tener sempre congiuntamente presenti i dati di stock e di flusso serve a contestualizzarli opportunamente e a ridurre la possibile strumentalità di comunicazioni basate su una sola misura[31]: ad esempio una forte e crescente incidenza sui flussi (come la quota di assunzioni a tempo determinato sul totale delle assunzioni) può accompagnarsi ad una incidenza debole e calante sullo stock (quota di contratti a tempo determinato sul totale dei contratti di lavoro in essere). E quindi l'indicazione di tendenza non può essere ricavata dal solo indicatore di flusso, ma basata su una contestuale valutazione dell'andamento dello stock corrispondente.

Per esplicitare tutti i problemi metodologici e documentare le soluzioni testate ed attivate, l'Osservatorio ha pubblicato[32], a partire dal 2016, una serie di report intitolati *Grammatica delle comunicazioni obbligatorie*. Originariamente erano previsti sei report, ne sono stati pubblicati quattro[33]; successivamente si sono aggiunti altri cinque report per aggiornamenti o per segmenti specifici (somministrazione, stagionalità).

Analizzare continuamente, in tutta la consistente produzione di monitoraggio della congiuntura, i flussi unitamente alle variazioni che essi determinano degli stock – variazioni sempre controllate anche confrontando altre fonti (Istat, Inps) – ha valorizzato e rassicurato la qualità dei risultati. Per questo l'ultima pagina dei report di monitoraggio congiunturale (*La Bussola*, *il Sestante*) ha sempre proposto il confronto con altre fonti (Istat o Inps), commentando le concordanze nell'identificazione delle tendenze in atto e facendo, delle eventuali dissonanze, punto di partenza per utili approfondimenti, spie da cogliere per capire meglio la realtà osservata e/o il punto di osservazione utilizzato.

A livello nazionale, invece, la valorizzazione statistica delle CO è proceduta nei primi anni con estrema lentezza, forse anche con qualche ragione seppur mai dichiarata (incertezze sul tasso di copertura dell'obbligo di comunicazione, in particolare relativamente alle cessazioni, con deficit rilevanti nelle regioni del Sud).

A forzare un cambio di passo sono state nell'inverno 2014-2015 la modifica della normativa sui licenziamenti varata con il *Jobs Act* (divenuta effettiva il 7 marzo) e la contestuale politica di incentivazione dei nuovi rapporti di lavoro a tempo indeterminato prevista dalla Legge di stabilità per il 2015 (L. n. 190/2014), attivata già da gennaio 2015. A quel punto era di grande interesse, per l'arena politica e per i media, conoscere l'impatto delle nuove disposizioni sui flussi di rapporti di lavoro. L'Osservatorio di Veneto Lavoro ha segnalato immediatamente il forte impatto sui rapporti di lavoro a tempo indeterminato – ancora estenuati dalla lunghissima crisi post 2008 – e ha pubblicato un primo report già il 4 marzo con dati aggiornati su tutto il bimestre gennaio-febbraio. Il 10 aprile entra in campo l'Inps con il nuovo Osservatorio Precariato, per iniziativa del neopresidente Tito Boeri, riportando nel primo numero le statistiche,

30 Con il passare degli anni gli archivi CO giungeranno a includere tutti gli occupati e allora si potrà stimare lo stock di posti di lavoro in essere giorno per giorno. Per ora il sistema ignora, per definizione, coloro che sono stati assunti prima del 2008 e sono ancora occupati: essi emergono al momento della comunicazione della cessazione, momento in cui viene registrata retrospettivamente anche la data dell'assunzione alimentando in tal modo i flussi antecedenti il 2008.

31 Come capita sovente sia in tanti articoli giornalistici sia nelle dichiarazioni di politici e sindacalisti.

32 https://www.venetolavoro.it/tempi-metodi

33 È interessante notare che non sono stati pubblicati (né redatti) i due report originariamente previsti e dedicati rispettivamente a *Il flusso delle informazioni CO. Cenni di storia. Modelli, canali di trasmissione. Criticità, metodi ed esperienze di validazione* (n. 1) e *Il frutto semiproibito: esperienze di integrazione con altri database* (n. 4). Non a caso due tematiche fin troppo sensibili: il primo, concernente la validazione (controllo) dei flussi informativi, implicava il confronto con la storia delle procedure di validazione e quindi con le connesse ispide paternità (essendo per ovvie ragioni tali procedure di interesse, oltre che dell'Osservatorio, anche dell'area informatica e dell'area operativa); il secondo implicava addentrarsi nel terreno minato della privacy.

basate sulle denunce mensili Uniemens, sui flussi di assunzione e trasformazione nei mesi di gennaio e febbraio. Ovviamente anche il Ministero, detentore delle CO, data la pressione mediatica crescente per avere informazioni su cosa sta succedendo nel mercato del lavoro, a quel punto non poteva più esimersi dal produrre qualche dato e infatti il 23 aprile pubblica la prima molto asciutta *Nota flash* intitolata *La dinamica dei contratti di lavoro. Prime evidenze ricavate dal Sistema Informativo delle Comunicazioni Obbligatorie.* L'aggiornamento dei dati è a marzo ed è corredato da un calendario delle successive uscite, con annuncio di un monitoraggio mensile continuo messo a disposizione di tutti.

Così per la prima volta nel giro di pochi mesi, diverse istituzioni mettono sul tavolo, annunciando l'avvio di una nuova prassi, dati tempestivi di flusso ricavati da fonti amministrative. I media, frettolosi, in buona parte ci capiscono poco, non sanno proprio come mettere in relazione i nuovi dati di flusso con le più familiari statistiche sullo stock di occupati (quelle tradizionali di fonte Istat), prendendo così delle vere e proprie cantonate, incapaci di distinguere novità da persistenze[34]. E, di fronte alla complessità derivante anche dall'inedito pluralismo delle fonti, si fanno guidare nell'interpretazione, come fosse bussola sicura, dal proprio posizionamento politico.

Per di più, il 25 agosto, il Ministero pubblica dati di flusso senza controllarne gli effetti sugli stock, ed è costretto a rettificarli rapidamente[35]. Successivamente il rilascio di dati CO a livello nazionale ritorna ad essere prudentemente rarefatto.

L'incidente è il pretesto per una diffusa richiesta – soprattutto di ambienti giornalistici e politici – di ridurre la presunta cacofonia statistica derivante dalla pluralità di metodi, di fonti e di soggetti produttori e divulgatori, per lasciar spazio a pochi dati, comprensibili intuitivamente ed esaurienti per definizione (il mito del numero semplice e giusto che parla da sé)[36].

Il Ministero risponde attivando un'iniziativa congiunta in partnership Ministero-Inps-Istat-Inail-Anpal con la *mission* (limitata) di condividere le informazioni e l'obbligo di raccordarsi. Gli enti lavorano anche formalmente insieme per alcuni anni dando vita a una produzione congiunta – di report trimestrali e annuali[37] – che ovviamente non partorisce l'impossibile "numero unico e definitivo" né sostituisce la produzione di ciascun ente ma si aggiunge, fornendo utili chiarimenti (metadati) e mostrando come le divergenze tra le fonti siano (quasi) sempre spiegabili e motivate dalle specificità di ciascuna fonte.

Le connessioni con i temi all'ordine del giorno: analisi congiunturali e ricerche ad hoc

La disponibilità di una base dati analitica come le CO ha consentito e consente di produrre dati ed elaborazioni con due diverse finalità: a. di analisi congiunturale, tanto più interessante quanto più tempestiva; b. di ricerca su temi specifici, rilevanti soprattutto se incrociano questioni particolarmente sentite nel dibattito pubblico e se fanno emergere problemi specifici, dando loro 'corpo' statistico.

Sotto il profilo congiunturale il lavoro sulle CO ha consentito di seguire da vicino l'evoluzione del mercato del lavoro, con un lag temporale rispetto agli eventi che, a partire dall'emergenza Covid, è sceso sotto il mese: dal 2020, infatti, è stato attivato un rilascio mensile di dati congiunturali (relativi al settore privato, escluse le agenzie di somministrazione) entro 15-20 giorni dalla fine del mese (in precedenza il lag temporale medio dei

34 Cfr. il titolo (in prima pagina) de *Il Fatto quotidiano* del 12 aprile 2015 (*Il miracolo del Jobs Act: 13 posti di lavoro*) e l'analogo, confuso ed erroneo, commento di Scalfari nell'editoriale de *La Repubblica* del medesimo giorno, entrambi riportati e commentati in Anastasia B., "Le statistiche (i numeri) aiutano a capire il mercato del lavoro o servono a confondere?", in *Economia e società regionale*, n. 3, 2017.

35 La nota attualmente disponibile nel sito è quella corretta https://www.lavoro.gov.it/temi-e-priorita/occupazione/Pagine/Studi-e-statistiche?page=5.

36 Ampia e interessante la ricostruzione di Francesco Nespoli di tutta "la battaglia di dati" che ha accompagnato il primo tempo del *Jobs Act*. L'autore utilizza questa vicenda per documentare un caso di fallimento della comunicazione politica (l'altro caso è la vicenda sindacale di Pomigliano); forse non sottolinea a sufficienza l'oggettiva difficoltà di molti dei protagonisti (giornalisti e politici) a maneggiare con qualche avvertenza e coerenza le (plurali) statistiche disponibili: "buttarla in politica", alla fine, si rivela come l'*escamotage* rassicurante e la soluzione ad ogni difficoltà interpretativa. Cfr. F. Nespoli (2018).

37 L' Accordo tra il Ministero del Lavoro e delle politiche sociali, l'Istat, l'Inps, l'Inail è stato firmato originariamente il 22 dicembre 2015 e prorogato fino al dicembre 2021. Nel 2017 si è aggiunta l'Anpal. La prima *Nota trimestrale sulle tendenze dell'occupazione*, relativa al terzo trimestre 2016, è stata diffusa il 28 dicembre 2016, l'ultima – relativa al quarto trimestre 2021 – è stata diffusa il 22 marzo 2022. Tra il 2017 e il 2020 sono stati pubblicati quattro rapporti annuali significativamente intitolati *Il mercato del lavoro. Una lettura integrata*. Tanto le *Note trimestrali* che i *Rapporti annuali* sono stati pubblicati contemporaneamente sui rispettivi siti web delle cinque istituzioni coinvolte. L'Accordo non è stato successivamente rinnovato, forse perché se ne ritenevano raggiunti gli obiettivi o forse perché nel frattempo – come si suol dire – "tutto è cambiato", soprattutto il Governo.

rilasci – solo trimestrali – era di circa un mese e mezzo)[38]. Teoricamente – avvertendone l'interesse[39] – si potrebbe arrivare a produrre un indicatore affidabile della dinamica delle assunzioni con un ritardo di un giorno e un indicatore di variazione dei posti di lavoro (tenendo conto quindi delle cessazioni) con un ritardo di una settimana. Queste attività di analisi congiunturale, proprio per la loro tempestività e ripetitività, implicano una attenta standardizzazione: gli indicatori da produrre devono essere ricorrenti e stabili.

Sotto il profilo dell'attività di ricerca, l'utilizzo in profondità dei dati CO ha consentito di fiancheggiare – portando solide evidenze a una discussione pubblica almeno potenzialmente più informata – diversi momenti del dibattito pubblico, in particolare a proposito di:

- contratti a termine: ragioni della loro diffusione, effetti delle politiche di contrasto ripetutamente attivate (Legge Fornero del 2012, *Jobs Act* del 2015, Decreto Dignità del 2018);
- politiche di incentivazione delle assunzioni con contratti a tempo indeterminato, in particolare l'esonero contributivo già citato previsto in Legge di bilancio 2015, monitorato fino alla sua conclusione;
- condizione giovanile: tempi di inserimento nel mercato del lavoro, tirocini ecc.;
- effettività delle normative sull'inserimento al lavoro dei disabili;
- lavoratori immigrati: dimensioni e caratteristiche della loro presenza nel sistema produttivo regionale, nesso con le politiche di ingresso (regolarizzazioni ecc.), specificità nell'interazione con le istituzioni del mercato del lavoro (il caso dei cinesi)[40].

Per la trattazione di molti di questi temi è stato straordinariamente utile il ricorso a dati longitudinali, via via sempre più riccamente disponibili man mano che ci si andava allontanando dall'anno zero delle CO (il 2008)[41].

La qualità delle analisi e delle statistiche è e resta condizionata dalla qualità dei dati che le imprese e i loro consulenti immettono nel sistema. Molte variabili potenzialmente di grande interesse non possono essere sfruttate a pieno perché frequenti sono gli omissis o vi è necessità di riclassificazioni/integrazioni: è il caso, ad esempio, dei titoli di studio, variabile di grande rilevanza, ma alquanto critica con riferimento soprattutto ai lavoratori immigrati (per i quali la difficoltà di comunicazione e codifica è ben nota). Per molte variabili l'integrazione con altre banche dati comporterebbe sicure possibilità di controllo/arricchimento/completamento delle informazioni, ma la collaborazione tra istituzioni diverse è frenata sia dalla loro scarsa propensione a collaborare (se non quando obbligate da norme) sia da oggettive difficoltà, anche tecnologiche (ogni amministrazione e quindi ogni banca dati ha un suo linguaggio, utilizza proprie definizioni e classificazioni ecc.) sia dall'incombente timore di una collisione con l'iceberg della privacy, che si innalza maestoso sul mare delle informazioni: fa poca paura ai transatlantici delle piattaforme e dei big data ma spaventa moltissimo le amministrazioni pubbliche.

E la disoccupazione?

Nell'immaginario pubblico i Centri per l'impiego ancor oggi – come prima gli Uffici di collocamento – sono i luoghi che hanno a che fare soprattutto con i disoccupati. Come mai allora l'attenzione, anche dell'Osservatorio di Veneto Lavoro, si è focalizzata soprattutto sui dati occupazionali desunti dalle CO? Perché la disoccupazione è passata in secondo piano, almeno come oggetto di ricerca?

Due risposte si possono dare, tra loro complementari:

a. l'utilizzo dei dati sui disoccupati a fini statistici implica un'attività consistente di manipolazione dei dati amministrativi di base. Tradizionalmente le statistiche sugli stock di disoccupati amministrativi – quando rese disponibili –

38 Nella fase iniziale del Covid, di fronte all'improvvisa radicale incertezza sull'impatto che il lockdown e le politiche straordinarie attivate (Cig, blocco dei licenziamenti ecc.) potevano avere sul mercato del lavoro, è stato attivato (con un primo report del 26 marzo 2020, *Emergenza Covid-19: l'impatto sul lavoro dipendente in veneto nel primo mese (23 febbraio-22 marzo 2020),* collana Misure n.88) un monitoraggio quindicinale, proseguito con 11 report fino a luglio. In quella fase l'Osservatorio ha collaborato anche con la Banca d'Italia: cfr. il report del 9 aprile 2020 *Prime indicazioni dal mercato del lavoro: i dati delle comunicazioni obbligatorie del Veneto*, a cura di B. Anastasia, G. Bovini (Banca d'Italia), M. Rasera (Veneto Lavoro), E. Viviano (Banca d'Italia) nella collana specifica della Banca d'Italia https://www.bancaditalia.it/pubblicazioni/note-covid-19/index.html.

39 Negli Stati Uniti la Borsa reagisce immediatamente alla comunicazione, il primo venerdì di ogni mese alle ore 8.30 locali, del dato sui *Non-Farm Payrolls* (occupati retribuiti non agricoli) che misura la variazione mensile del numero di occupati dipendenti sulla base della *Current Employment Statistics Survey*, che coinvolge circa 150.000 imprese. Se avessero le CO…

40 Per i singoli lavori di ricerca su questi temi si vedano soprattutto le due collane *Misure* (126 numeri dal 2008) e *Tartufi* (63 numeri dal 2000) nel sito di Veneto Lavoro.

41 È stato identificato e testato anche un sistema di indicatori standard di tipo longitudinale, che però non ha dato luogo ad una produzione ricorrente: si veda Anastasia *et al.* (2018).

hanno sempre avuto il sapore dell'irrealtà: non a caso non venivano – e non vengono – citate nel dibattito pubblico. Le ragioni di ciò sono cambiate nel corso del tempo, man mano che è cambiata la regolazione del mercato del lavoro. Fino al 1991 (L. n. 223) – quando venne meno anche formalmente il sistema della chiamata numerica, già in crisi di fatto – essere certificati amministrativamente come disoccupati era un passaggio imprescindibile antecedente all'assunzione: non a caso le due nozioni di 'avviato' e 'assunto' si sovrapponevano, i due termini erano intesi come sinonimi, anche se tali non erano. Inoltre, l'anzianità di disoccupazione conferiva molti vantaggi (priorità negli avviamenti ecc.; alcuni di questi – come l'esonero dai ticket sanitari – sono sopravvissuti a lungo, forse sopravvivono ancor oggi). Di conseguenza il dato amministrativo sopravalutava enormemente la dimensione dei disoccupati effettivamente alla ricerca di lavoro. Successivamente anche la regolazione amministrativa ha cercato di approssimare una definizione di disoccupato più vicina o quantomeno confrontabile con la nozione utilizzata nella statistica ufficiale. Ma le divergenze permangono rilevanti. Dal lato degli ingressi – identificati con il rilascio della dichiarazione di disponibilità – di fatto non tutti i disoccupati si rivolgono ai CPI: lo devono fare certamente solo coloro che accedono a sostegni pubblici (in primis la NASpI). Ma è ancor più complesso il trattamento delle informazioni sulle uscite: non tanto quelle parziali, per lavoro temporaneo, quanto quelle definitive per transizione all'inattività o al lavoro autonomo (mentre quelle al lavoro dipendente sono agevolmente documentate dalle CO). Ne derivano stock sovradimensionati, che periodicamente devono essere ripuliti e che rendono assai arduo il lavoro di costruzione di serie storiche affidabili[42];

b. i disoccupati costituiscono il 'core business' dei CPI e delle relative strutture regionali. Sono la platea designata per gli interventi di politica attiva: orientamento, accompagnamento, formazione ecc. Documentare il successo o l'insuccesso delle politiche ad essi rivolte significa interferire con il racconto – e l'implicita valutazione – che le amministrazioni interessate (Stato, Regioni) propongono circa le proprie iniziative e i risultati conseguiti.

Scontando questi problemi, l'Osservatorio di Veneto Lavoro, soprattutto dopo il *Jobs Act* e la riforma dei Centri per l'impiego ivi abbozzata, ha comunque ripetutamente descritto consistenza e caratteristiche dei disoccupati amministrativi in Veneto[43]. Tale attività – a differenza di quella rivolta alla valorizzazione delle CO – non ha dato luogo ad aggiornamenti costanti e ripetuti secondo format stabili.

Su questo argomento – dimensioni e caratteristiche effettive della disoccupazione e impatti delle politiche attive – ci sono ampi spazi di approfondimento, in particolare per capire (dimensionare) i problemi che stanno attorno all'incontro domanda-offerta (mismatch, canali di intermediazione, incentivi e ruolo degli ammortizzatori, adeguatezza dei profili professionali, impatto della formazione, salari d'ingresso ecc.). Tanto più che nel tempo si è profondamente modificato anche lo spazio d'azione dei CPI, che – almeno in Veneto, come in alcune altre regioni – hanno innovato metodologie di lavoro e modalità di presenza, da ultimo grazie anche agli incentivi in tal senso legati ai programmi europei Garanzia Giovani prima e PNRR-GOL dopo. Perciò non ha senso – come a volte si fa ancora – riproporre indicatori fuorvianti, strampalati in radice (come ad es. la percentuale di assunti 'per merito' dei CPI sul totale delle assunzioni), per mostrare che "i CPI non funzionano"[44]. Serve spostare l'attenzione su platee effettivamente trattate, costi sostenuti, obiettivi praticabili e azioni effettive per raggiungerli, valorizzando la massa di informazioni anche qualitative oggi disponibili in misura nettamente superiore al passato: l'ovvio riferimento è alle SAP (Schede anagrafico-professionali), una delle colonne portanti del SIU (Sistema informativo unitario delle politiche del lavoro), l'infrastruttura nazionale pensata per costituire un'unica base informativa condivisa tra Stato, Regioni e Servizi per l'impiego a supporto sia delle politiche attive sia della produzione statistica. Nato formalmente con il D.Lgs. n. 150/2015 (*Jobs Act*),il SIU è un'architettura di interoperabilità che collega e armonizza archivi già esistenti (SAP, CO, dati dei Centri per

42 La relazione tra disoccupati amministrativi (CPI) e disoccupati ufficiali (Istat) è analoga a quella, ancor più complessa, tra poveri amministrativi (oggetto delle politiche di Reddito di cittadinanza, attualmente col brand rinnovato di Assegno di inclusione) e poveri secondo i dati ufficiali (Istat). Si ragiona e si dibatte con i numeri Istat, si agisce con i numeri dell'Inps. Ciò complica di molto la valutazione degli effetti delle politiche attivate al riguardo.

43 Cfr. il prodotto più recente al riguardo Gambuzza *et al.* (2022).

44 Su questo punto – in particolare sull'indeterminatezza della mission dei Centri per l'impiego e sul ricorso a indicatori fallaci in materia di intermediazione – cfr. l'excursus proposto in Anastasia e Santoro (2018).

l'impiego, informazioni su NASpI, DIS-COLL e altre misure)[45]. I dati quindi ci sarebbero, ricchi e longitudinali (anagrafica, storia lavorativa, prese in carico, politiche attive): ad oggi l'uso del SIU a fini conoscitivi risulta limitato, di fatto non è (ancora) una fonte statistica 'matura', l'accesso ai dati resta finalizzato essenzialmente alle attività operative ed è poco (pochissimo) aperto alla comunità scientifica. Il risultato è una sorta di paradosso informativo: il SIU è stato pensato anche per monitoraggio e valutazione ma di fatto la produzione statistica resta ancillare, strumentale alla gestione dei programmi, non autonoma come sistema informativo pubblico. Ciò è il frutto di diversi fattori: della governance instabile (vicenda Anpal ecc.) delle politiche attive negli ultimi 10-15 anni; della forte frammentazione regionale dei servizi; delle priorità amministrative (adempimenti, pagamenti, condizionalità) rispetto a quelle conoscitive; dell'assenza di un mandato chiaro e continuativo di *statistical reporting* sul lato politiche attive. Quando sarà utilizzato in modo sistematico il SIU potrà produrre conoscenza utile sia sui percorsi dei disoccupati sia sugli aspetti qualificanti dei CPI come organizzazioni operative: frequenza e tipologia dei servizi erogati, distinzione tra utenti solo formalmente presi in carico, utenti seguiti, utenti intensivamente trattati. Oggi questa distinzione è opaca, anche per i decisori pubblici. Grazie all'integrazione con il SIISL si possono (si potrebbero) confrontare beneficiari e non beneficiari di specifiche misure di politica attiva (incentivi, formazione ecc.) nonché monitorare accuratamente i percettori di indennità di disoccupazione (NASpI ecc.) e i loro percorsi. Producendo sistematicamente statistiche sul funzionamento delle politiche del lavoro, il SIU renderebbe meglio osservabile ciò che oggi risulta illuminato piuttosto episodicamente e parzialmente. Di recente, anche per dar conto delle attività connesse al PNRR (Piano nazionale di ripresa e resilienza) e al Programma specifico GOL (Garanzia occupabilità lavoratori), l'Osservatorio di Veneto Lavoro ha prodotto un'ampia ricerca sugli utenti dei Centri per l'impiego, proponendone un'originale classificazione, e sulle attività ad essi dedicate. Si tratta di analisi prodromiche anche alla valutazione dell'effettività della condizionalità (Bertazzon *et al.* 2026).

Sperimentazioni di link con altre basi-dati: lo studio della crisi del 2008 con i dati sui redditi fiscali (Irpef)

L'integrazione tra banche dati di amministrazione diverse aumenta notevolmente sia la qualità dei dati (perché permette controlli/integrazioni) sia la possibilità di approfondire la conoscenza dei fenomeni osservati. Per le ragioni già ricordate è attività irta di ostacoli, in genere trascurata anche per lo scoraggiamento provocato dagli elevati costi che si incontrano in tale percorso (in termini di tempo, procedure, difficoltà burocratiche ecc.).

Merita segnalare un caso di relativo – perché episodico – successo: quello dell'integrazione, per il Veneto, tra microdati fiscali (dichiarazioni Irpef della Regione Veneto) e dati CO – che ha originato il database MelFisco – con l'obiettivo di arricchire l'analisi dei movimenti occupazionali con l'informazione sulla dimensione salariale, per studiare l'impatto della lunga fase contrassegnata dalla doppia crisi 2008-2014 (Anastasia *et al.* 2012; Anastasia *et al.* 2020). L'utilizzo congiunto delle due fonti ha consentito di identificare transizioni tra lavoro dipendente e lavoro indipendente (e viceversa) nonché nessi tra variazioni del reddito e mobilità dei lavoratori. Con ogni probabilità si tratta di uno dei pochi esempi italiani in cui il *linkage* tra archivi fiscali e del lavoro è stato utilizzato per fini conoscitivi. Dal punto di vista istituzionale si è dimostrata la fattibilità tecnica e analitica dell'uso di microdati fiscali a livello regionale, superando il loro tradizionale confinamento all'ambito nazionale.

La messa a disposizione dei ricercatori dei microdati e il Public Use file

Un'istituzione proprietaria di dati di potenziale interesse pubblico può metterli a disposizione del pubblico scegliendo tra soluzioni molto diverse per ricchezza di informazioni fornite e trasparenza:

a. produrre poche o pochissime tabelle essenziali, più o meno appropriate e costruite con indicatori più meno di immediata significatività;
b. consentire l'accesso tramite il proprio sito web a dati ampiamente e facilmente navigabili, con la possibilità per l'utente di incrociare a sua scelta un ampio numero di variabili con disaggregazioni anche molto dettagliate[46];

45 A sua volta il SIU è parte integrante del SIISL (Sistema informativo per l'Inclusione sociale e lavorativa), contenitore istituito (L. n. 85/2023) presso il Ministero del Lavoro e realizzato dall'Inps, nato per gestire Assegno di Inclusione (ADI) e Supporto formazione e lavoro (SFL).

46 L'esperienza nazionale più consolidata e ampia in materia è quella degli *Osservatori Inps* che spaziano su tante dimensioni della vita socioeconomica: dalle pensioni ai flussi di assunzioni e cessazioni; dalle dichiarazioni delle condizioni economiche della famiglia agli ammortizzatori e agli incentivi alle imprese per assunzioni ecc. Si tratta di un'esperienza sostanzialmente isolata: non è stata seguita in modo altrettanto sistematico e continuo da altre grandi amministrazioni nazionali (come del resto si può dire, a livello regionale, per l'esperienza di Veneto Lavoro). Per di più attualmente l'esperienza Inps degli *Osservatori* è in fase di malaugurato ridimensionamento/ripensamento per la debordante ingerenza delle prescrizioni della privacy.

c. fornire, su richiesta di soggetti abilitati (ricercatori ecc.), basi di microdati integrali (anonimizzati e manipolati al minimo) in modo che anche ricercatori esterni possano replicare/controllare i risultati ottenuti dagli analisti interni dell'amministrazione e ricercarne ulteriori secondo altri obiettivi e interessi di ricerca.

L'Osservatorio di Veneto Lavoro ha praticato tutte queste vie.

Per la produzione ricorrente ha messo e mette a disposizione sul proprio sito, con accesso libero, con grande tempestività dati molto disaggregati – arrotondati a 5 per ottemperare le prescrizioni in materia di privacy – relativi sia a tutti i flussi CO (e relativi saldi) sia alle principali dimensioni della disoccupazione amministrativa[47].

Per quanto ricchi, questi dati navigabili consentono l'esplorazione di un numero limitato di variabili: per ricerche in profondità e controllare/replicare quanto prodotto dall'Osservatorio serve disporre di microdati liberamente elaborabili. A questo scopo è stato costruito e aggiornato nel tempo un *Public Use File* (PUF), basato su un vasto insieme di dati amministrativi detenuti da Veneto Lavoro, che viene rilasciato su richiesta per esigenze motivate di ricerca[48]. L'esperienza è stata ufficializzata nel 2004, con il primo rilascio. Il primo PUF[49] costruito è stato denominato Giove. In seguito all'aumento della profondità temporale della base dati, della qualità dei controlli e delle esigenze di ricerca, Giove è stato ripetutamente ristrutturato, cambiando nome prima in Planet, poi da ultimo in Mercurio. Queste edizioni successive del PUF originario incorporano avanzamenti metodologici nella pulizia dei dati, nella ricostruzione delle storie individuali e nell'anonimizzazione statistica. Non si tratta solo di cambi di nome, ma di vere *re-ingegnerizzazioni* del dato, per renderlo meglio utilizzabile a scopi scientifici (e senza compromettere la necessaria riservatezza). Il PUF rappresenta un modello di trasparenza nell'utilizzo dei dati amministrativi per finalità conoscitive, concepiti dunque come bene pubblico, non come patrimonio, di fatto semi-privatistico, degli enti proprietari. Esso è stato richiesto e utilizzato[50] da oltre un centinaio di ricercatori, soprattutto universitari, attivi anche in strutture all'estero. Infine, si può senz'altro segnalare che auto-obbligarsi a produrre il PUF responsabilizza la struttura che lo rilascia, ne consolida la credibilità e genera, di rimbalzo, un aumento della qualità dei dati e delle elaborazioni.

4. Per un buon uso dei dati: dare a Cesare quel che è di Cesare

Anche grazie ai dati amministrativi, oggi sappiamo molto di più sul mercato del lavoro rispetto al passato. Possiamo seguirne l'evoluzione con tempestività, indagarne aspetti specifici sotto il profilo territoriale e/o settoriale, leggerli in longitudinale, collegarli agli incentivi e alle politiche attive. Lo sfruttamento di tutte queste potenzialità può ancora migliorare per arricchire la qualità delle conoscenze disponibili.

Parallelamente il conflitto mediatico sull'interpretazione dei numeri rimane acceso. Del resto quanto più i dati sono abbondanti, tanto più si allargano le opzioni disponibili – quali numeri valorizzare, quali ignorare, quale periodo scegliere come riferimento, a quali indicatori dar maggiore peso ecc. – spesso guidate dall'obiettivo di avvalorare narrazioni semplici e stilizzate, per confermare la bontà delle proprie posizioni (nel caso dell'utilizzo nell'arena politica) o per indicare novità mirabolanti, in genere negative (nel caso dell'utilizzo nell'arena mediatica).

Il possibile miglioramento continuo della conoscenza dei fenomeni economico-sociali, grazie anche ai dati amministrativi, non può proporsi l'obiettivo di ridurre il conflitto interpretativo sullo stato delle cose (ad es. se la flessibilità contrattuale sia troppa o troppo poca) né sui desiderata da indicare/proporre (quale livello di salario minimo sia accettabile ecc.). Può però fornire/imporre un linguaggio (definizioni ecc.) e misure scientificamente fondate e trasparenti, e per questo condivisibili, a supporto di una descrizione sufficientemente obiettiva della realtà e delle tendenze in corso, riducendo i margini per strumentalizzazioni palesemente fuorvianti, circoscrivendo lo spazio quindi alle vere (e legittime) differenze tra gli stakeholder che devono riguardare non tanto il "com'è?, come va?" quanto il "che fare?" "che obiettivi vogliamo perseguire?".

La produzione statistica deve servire a depotenziare gli imperativi narrativi della

47 Cfr. https://www.venetolavoro.it/numeri.

48 Cfr. https://www.venetolavoro.it/public-use-file. L'ultima release attualmente disponibile (Mercurio 4.0) è aggiornata a dicembre 2022.

49 L'esperienza del PUF, che riproduce a uso esterno la medesima banca di microdati a disposizione dei ricercatori interni, non va confusa con quella degli open-file (in genere dati statistici scarsamente sistematici, non micro-dati) che si possono trovare in molti siti di amministrazioni pubbliche.

50 Cfr. ad esempio uno dei primi studi scientifici sull'impatto nel 2015 dell'esonero contributivo e del *Jobs Act*, condotto sui microdati CO del Veneto: Sestito e Viviano (2016).

comunicazione politico-mediatica, evitando che i numeri diventino solo il loro ornamento e che, per reazione, la ricerca più avanzata si rifugi in attività accademiche autoreferenziali.

Bibliografia

Agenzia per l'impiego del Veneto (2000), *Solo una grande giostra? La diffusione del lavoro a tempo determinato*, Milano, Franco Angeli

Anastasia B. (1993), Una merce maltrattata: il dato. Con un decalogo per il consumatore, *Oltre il ponte. Economia e Società regionale*, n.43-44, pp.186-195

Anastasia B., Basso S. (2015), *Sei anni (2009-2014) di Cassa integrazione in deroga in Veneto: quanto ha contato, come ha funzionato, che esiti ha prodotto*, I Tartufi n.44, Venezia-Mestre, Veneto Lavoro

Anastasia B., Bragato S., Gambuzza M., Rasera M. (1996), La varietà dei flussi verso i lavori. Un'analisi di microarea, in Agenzia per l'impiego del Veneto (a cura di), *Il mercato del lavoro nel Veneto. Tendenze e politiche. Rapporto 1996*, Milano, Franco Angeli

Anastasia B., Disarò M., Gambuzza M., Rasera M. (2009), *Comunicazioni obbligatorie e analisi congiunturale del mercato del lavoro: evoluzione, problemi metodologici, risultati*, I Tartufi n.35, Venezia-Mestre, Veneto Lavoro

Anastasia B., Emireni G., Giommoni T. (2020), *Siamo nella stessa tempesta ma non sulla stessa barca. L'impatto della "grande recessione" sui redditi dichiarati delle persone fisiche: contributi da un'analisi longitudinale*, I Tartufi n.50, Venezia-Mestre, Veneto Lavoro

Anastasia B., Emireni G., Vanin F. (2012), *Il mercato del lavoro in Veneto… visto dalle dichiarazioni dei redditi Irpef*, I Tartufi n.40, Venezia-Mestre, Veneto Lavoro

Anastasia B., Gambuzza M., Gatti F., Maschio S., Rasera M. (2018), *L'apporto conoscitivo degli indicatori longitudinali per l'analisi del mercato del lavoro*, Veneto Lavoro

Anastasia B., Maschio S. (2017), *La qualità dei dati SILV sull'orario di lavoro*, Veneto Lavoro

Anastasia B., Santoro G. (2018), Jobs Act: politiche attive e ammortizzatori sociali, *Economia italiana*, n.2-3, pp.37-70

Anastasia B., Trivellato U. (2001), Servizi pubblici all'impiego, sistemi informativi e ricadute di conoscenza: esperienze pilota con Netlabor, in Trivellato U. (a cura di), *Servizi per l'impiego e ricerche sul lavoro. L'esperienza del Veneto*, Quaderni di economia del lavoro n.72, Milano, Franco Angeli, Cap.2

Bassi F., Rasera M., Gambuzza M. (2001), Il sistema informatizzato Netlabor: struttura, caratteristiche e qualità, in Trivellato U. (a cura di), *Servizi per l'impiego e ricerche sul lavoro. L'esperienza del Veneto*, Quaderni di economia del lavoro n.72, Milano, Franco Angeli, Cap.3

Bertazzon L., De Colle M., Maschio S., Nadalin F., Rocco I. (2026), *Monitoraggio dei disoccupati utenti dei Centri per l'impiego in Veneto*, I Tartufi n.65, Venezia-Mestre, Veneto Lavoro

Capone L. (2025), Landini dà i numeri sul fiscal drag, Editoriale, *Il Foglio*, 21 ottobre

Contini B., Revelli R. (1992), *Imprese, occupazione e retribuzioni al microscopio: studi sull'economia italiana alla luce delle fonti statistiche INPS*, Bologna, il Mulino

Contini B., Trivellato U. (2005), *Eppur si muove. Dinamiche e persistenze del mercato del lavoro italiano*, Bologna, il Mulino

Favero G., Trivellato U. (2000), Il lavoro attraverso gli 'Annali': dalle preoccupazioni sociali alla misura della partecipazione e dei comportamenti nel mercato del lavoro, in Geretto P. (a cura di), *Statistica ufficiale e storia d'Italia. Gli "Annali di Statistica" dal 1871 al 1997*, Annali di Statistica, Serie X, vol. 21, Roma, Istat, pp.223-302

Gambuzza M., Gatti F., Maschio S., Rocco I., Rasera M. (2022), *I disoccupati amministrativi dei Centri per l'impiego*, I Tartufi n.55, Venezia-Mestre, Veneto Lavoro

Istat (2025), *Pratiche per il trattamento delle informazioni nel processo di produzione statistica: compliance e metodi*, Roma, Istat

Ministero del Lavoro e della previdenza sociale (2006), Gruppo di lavoro per il monitoraggio degli interventi di politica occupazionale e del lavoro, *Aggiornamento del quadro informativo sulle politiche del lavoro*, Roma, Ministero del Lavoro e della previdenza sociale

Nespoli F. (2018), *Fondata sul lavoro. La comunicazione politica e sindacale del lavoro che cambia*, Modena, Adapt University Press

Nespoli F. (2025), Quella del racconto dei dati sul lavoro è una battaglia infinita (e anche inutile), *Bollettino Adapt*, n.43, 8 dicembre

Sestito P., Trivellato U. (2010), *Indagini dirette e fonti amministrative: dall'alternativa all'ancora incompiuta integrazione*, Working Paper, Padova, Università di Padova, Dipartimento di Scienze Statistiche

Sestito P., Viviano E. (2016), *Hiring incentives and/or firing cost reduction? Evaluating the impact of the 2015 policies on the Italian labour market*, Questioni di economia e finanza, Occasional Papers n.325, Roma, Banca d'Italia

Bruno Anastasia
brunoanasta@gmail.com
Si occupa di analisi del mercato del lavoro. Ha diretto fino al 2019 l'Osservatorio sul mercato del lavoro regionale di Veneto Lavoro e ha collaborato con numerose istituzioni regionali e nazionali, tra cui Inps, Ministero del Lavoro, CNEL, Fondazione Nord Est e la Commissione d'indagine sul lavoro promossa da CNEL, Camera dei Deputati e Senato. Dal 2001 al 2006 è stato Presidente dell'Ires Veneto. Dal 1994 al 2001 è stato presidente del Coses di Venezia.

Perché essere interoperabili

Ontologie, vocabolari controllati e schemi di dati del Catalogo nazionale dati

Alessio Buonocore
ISTAT

Eugenio Concetti
ISTAT

Giampiero D'Alessandro
ISTAT

Adriana Fidanza
ISTAT

Stefania Pantoni
ISTAT

Roberta Radini
ISTAT

I dati amministrativi sono una risorsa strategica per supportare la Pubblica amministrazione nelle decisioni e nelle politiche basate su evidenze, come previsto dalla *European Data Strategy*. La valorizzazione di questi dati – attraverso qualità, condivisione e interoperabilità – è fondamentale per una gestione più efficiente e coordinata. Strumenti come ontologie, vocabolari controllati e schemi di dati, promossi dal Catalogo nazionale dati Schema.gov.it di cui Istat è soggetto attuatore, rafforzano l'interoperabilità semantica e migliorano la capacità della PA di utilizzare i dati in modo integrato e coerente.

Why be interoperable? Ontologies, controlled vocabularies and data schemas of the National Data Catalog

Administrative data constitute a key strategic asset for enabling evidence-based policy formulation and decision-making across the public sector, in line with the objectives of the European Data Strategy. Unlocking the full value of these data through enhanced quality, sharing mechanisms, and interoperability is crucial for fostering more efficient, coordinated, and data-driven public governance. Semantic assets – including ontologies, controlled vocabularies, and data models – made available through Schema.gov.it, the National Data Catalog, with ISTAT acting as the implementing body, play a pivotal role in strengthening semantic interoperability and improving the capacity of public administrations to integrate, exchange, and reuse data consistently across organizational and policy domains.

DOI: 10.53223/Sinappsi_2026-02-3

Citation
Buonocore A., Concetti E., D'Alessandro G., Fidanza A., Pantoni S., Radini R. (2026), Perché essere interoperabili. Ontologie, vocabolari controllati e schemi di dati del Catalogo nazionale dati, *Sinappsi*, XVI, n.2, pp.49-64

Parole chiave
Catalogo Nazionale Dati
Dati amministrativi
Interoperabilità

Keywords
National Data Catalog
Administrative data
Interoperability

Introduzione[1]

Nel contesto della trasformazione digitale delle pubbliche amministrazioni, i dati amministrativi rappresentano una risorsa strategica di primaria importanza per il policy making. Nati come strumento operativo per la gestione dei procedimenti amministrativi, costituiscono oggi un patrimonio informativo di enorme valore pubblico, in grado di supportare analisi statistiche, valutazioni di impatto e progettazione di politiche basate su evidenze. La loro natura di bene comune, generato grazie a risorse collettive e riferito a fenomeni di interesse

1 Il presente lavoro va considerato frutto del lavoro complessivo degli Autori. Tuttavia, il par. 1 va attribuito ad Alessio Buonocore, il par. 2 a Stefania Pantoni, il par. 3 ad Adriana Fidanza, il par. 4 ad Eugenio Concetti ed il par. 5 a Giampiero D'Alessandro; l'introduzione e le conclusioni a Roberta Radini. Si precisa che quanto qui scritto non corrisponde necessariamente alle posizioni espresse dall'Istituzione di appartenenza degli Autori.

generale, impone la necessità di una gestione orientata alla qualità, alla condivisione e al riuso. In questa prospettiva si colloca la *European Data Strategy* (Commissione europea 2020), che mira alla creazione di un mercato unico dei dati fondato sulla libera circolazione delle informazioni. Tale strategia promuove la cultura del dato come infrastruttura di conoscenza, valorizzando i principi FAIR – *Findable, Accessible, Interoperable, Reusable* – (Wilkinson *et al.* 2016) quali criteri fondamentali per garantire l'effettiva utilizzabilità dei dati pubblici e la loro integrazione in ecosistemi digitali comuni. Tuttavia, affinché i dati amministrativi possano diventare una leva reale per la conoscenza e per decisioni pubbliche più efficaci, è necessario che essi siano interoperabili e standardizzati. L'interoperabilità, nelle sue dimensioni tecnica, semantica, organizzativa e giuridica, costituisce infatti il prerequisito essenziale per trasformare l'attuale frammentazione dei sistemi informativi pubblici in una rete coerente e integrata di dati.

L'obiettivo di questo contributo è illustrare come strumenti quali ontologie, vocabolari controllati e schemi di dati pubblici rappresentino elementi chiave per realizzare l'interoperabilità semantica, favorendo la convergenza dei linguaggi e delle strutture informative delle amministrazioni. In particolare, verrà dedicata un'attenzione specifica a *schema.gov.it*, il Catalogo nazionale per la semantica dei dati - *National Data Catalog* (NDC), di cui l'Istat (Istituto nazionale di statistica) è soggetto attuatore. Il Catalogo costituisce oggi un riferimento centrale per la standardizzazione e la condivisione dei modelli dell'informazione all'interno della Pubblica amministrazione (PA) italiana.

Nella prima parte il contributo affronta il tema dei dati amministrativi come bene pubblico e infrastruttura di conoscenza. Viene poi approfondito il concetto di interoperabilità come requisito fondamentale per la valorizzazione dei dati, analizzandone i diversi livelli e i rischi connessi alla frammentazione dei sistemi informativi. Infine, si presentano gli strumenti semantici e gli standard nazionali ed europei che possono abilitare un ecosistema informativo realmente integrato e orientato al valore pubblico del dato, analizzando il caso di studio del Catalogo nazionale e gli impatti attesi per le politiche pubbliche.

1. I dati amministrativi come bene pubblico e infrastruttura di conoscenza

Con la locuzione 'dati amministrativi' si indicano le informazioni raccolte e conservate dalle pubbliche amministrazioni nello svolgimento delle loro funzioni istituzionali e gestionali (Amico *et al.* 2018; Istat 2021). Si tratta di dati generati primariamente per finalità operative (erogazione di servizi, gestione di registri e/o applicazione di norme ecc.) e non con scopi statistici o di ricerca. Rientrano nella categoria, ad esempio, i registri anagrafici e dello stato civile, gli archivi fiscali e previdenziali, i dati sanitari raccolti dalle aziende ospedaliere, i registri scolastici e universitari, le banche dati relative al lavoro e alle imprese.

È opportuno distinguere, in questa prospettiva, tra gestione dei dati per finalità amministrative e per finalità statistiche. Nel primo caso, il dato è prodotto e trattato nell'esercizio della funzione amministrativa: serve a istruire procedimenti, adottare atti, alimentare registri, erogare servizi digitali ed *e-service*, secondo esigenze di certezza giuridica, tracciabilità, semplificazione e buon andamento dell'azione pubblica (Spasiano 2017; Galetta 2021; Alberti 2022). Nel secondo caso, invece, il dato amministrativo è assunto come input di un processo statistico: viene documentato, selezionato, validato, integrato e trasformato secondo criteri metodologici orientati alla produzione di informazione aggregata, comparabile e riutilizzabile (UNECE 2011; Istat 2021; Ceccarelli 2025). Ne deriva una diversa accezione del concetto di fonte: in ambito amministrativo la fonte coincide con il sistema, il registro o il procedimento che genera il dato nell'esercizio di una competenza istituzionale; in ambito statistico, la fonte è invece valutata in rapporto alla sua qualità, copertura, coerenza, tempestività, stabilità definitoria e adeguatezza rispetto agli obiettivi conoscitivi (Wallgren e Wallgren 2007; Eurostat 2019). Lo stesso archivio amministrativo può quindi essere, al tempo stesso, strumento dell'azione amministrativa e fonte statistica, purché sia accompagnato da metadati, regole di qualità e processi di armonizzazione semantica che ne rendano possibile il riuso controllato.

Dunque, le informazioni nate per finalità operativo-amministrative hanno un valore notevole, che va oltre l'uso per il quale sono state prodotte: possono essere riutilizzate come fonte di conoscenza per analisi statistiche, valutazioni e progettazione di politiche pubbliche. In questi termini, i dati amministrativi vanno considerati non solo come strumento operativo, ma come una risorsa strategica sia per l'azione pubblica che per la comunità scientifica, quindi per la collettività intera. La loro natura li rende una forma di bene pubblico: la raccolta è finanziata da risorse collettive, il contenuto descrive fenomeni sociali, economici e territoriali di interesse generale e, come si diceva poc'anzi, il loro potenziale utilizzo va oltre le finalità primarie per le quali vengono generati, per tale motivo sono noti anche come 'dati secondari'.

I dati amministrativi offrono diversi vantaggi rispetto, ad esempio, alle rilevazioni dirette. Un primo

elemento distintivo è la copertura: diversamente dalle indagini campionarie, i dati amministrativi, di sovente, riguardano l'intera popolazione di riferimento, permettendo analisi dettagliate anche a livello micro e su scala locale; tale caratteristica aumenta la potenza analitica e consente di indagare fenomeni altrimenti difficilmente osservabili. Poiché i dati sono generati per usi amministrativi concreti, devono rispettare criteri di puntualità e precisione. Le linee guida UNECE (*United Nations Economic Commission for Europe*) sottolineano come la qualità dipenda dal rispetto di procedure standard nella raccolta e dalla tempestività degli aggiornamenti (UNECE 2021). I dati amministrativi possono, dunque, vantare un elevato livello di accuratezza e aggiornamento. Infine, i costi della loro raccolta e manutenzione sono già sostenuti dalle amministrazioni che li hanno prodotti, ne consegue che l'utilizzo dei dati amministrativi per fini statistici o di ricerca risponde pienamente al principio di sostenibilità economica.

Nonostante le potenzialità, l'effettiva valorizzazione dei dati amministrativi è ostacolata da diversi fattori. I sistemi informativi amministrativi sono spesso sviluppati autonomamente da enti diversi, con differenze nei modelli concettuali, nei metadati, e nelle definizioni adottate. Questa eterogeneità delle fonti ostacola l'integrazione e l'armonizzazione dei dati (UNECE 2011). Le linee guida dell'ESS (*European Statistical System*) evidenziano che la mancanza di armonizzazione concettuale e di metadati condivisi tra fonti diverse può compromettere la coerenza e la comparabilità dei dati (Eurostat 2019). Inoltre, l'uso dei dati raccolti per altre finalità comporta sfide legate alla protezione dei dati personali e sensibili; occorre bilanciare, dunque, la protezione dei dati personali con l'interesse pubblico alla conoscenza (GDPR -Regolamento UE 2016/679; Commissione europea 2017). In Italia, il Codice dell'amministrazione digitale (CAD) (D. Lgs. n. 82/2005) disciplina l'accesso telematico ai dati e al loro riutilizzo. Le evoluzioni del CAD hanno progressivamente rafforzato il principio della condivisione e interoperabilità dei dati tra PA, riconoscendo il patrimonio informativo pubblico come infrastruttura strategica per l'innovazione e l'efficacia dell'azione amministrativa. In particolare, il CAD introduce il principio della 'disponibilità dei dati' (art. 50[2]) secondo cui le pubbliche amministrazioni devono rendere accessibili alle altre amministrazioni i dati di cui sono titolari, nel rispetto delle normative vigenti in materia di protezione dei dati personali e sicurezza informatica (art. 51).

In questa prospettiva si inseriscono anche le basi di dati di interesse nazionale e, più recentemente, la Piattaforma digitale nazionale Dati, strumenti finalizzati a favorire l'interoperabilità applicativa e semantica tra sistemi informativi pubblici, favorendo l'integrazione dei sistemi informativi delle amministrazioni pubbliche e riducendo la duplicazione nella raccolta delle informazioni (art. 50-ter, co. 2).

L'evoluzione del CAD evidenzia dunque il passaggio da una concezione del dato amministrativo come patrimonio settoriale del singolo ente a una visione del dato quale risorsa condivisa, riutilizzabile e funzionale alla costruzione di un ecosistema informativo pubblico integrato.

In questo quadro, l'idea di dati amministrativi come bene pubblico sottolinea la necessità di politiche di governance condivise, in grado di garantire non solo la disponibilità ma anche la fruibilità, la qualità, e la sicurezza dei dati. L'adozione di standard comuni e di strumenti semantici (si veda par. 3) è essenziale per trasformare insiemi eterogenei di registri di dati amministrativi in una vera e propria infrastruttura di conoscenza a supporto delle decisioni pubbliche (si veda par. 5), così come delineato nei principi fondamentali espressi nella *European Data Strategy* (Commissione europea 2020). Il mercato unico che questa strategia punta a creare è volto alla condivisione dei dati tra i settori e gli Stati membri, promuovendo l'uso dei dati per l'innovazione e la crescita, nel rispetto delle normative europee. Questi obiettivi sono raggiungibili soltanto se si realizza l'interoperabilità dei dati (si veda par. 2).

2. L'interoperabilità come requisito fondamentale

Attualmente le PPAA producono, raccolgono e gestiscono grandi quantità di dati in formati differenti, con difformi metodi di gestione, replicandoli in diversi *repository* e pubblicandoli su portali che erogano servizi pubblici, senza armonizzazione di contenuto e di presentazione. Nell'ultima edizione del Censimento permanente delle Istituzioni pubbliche (Istat 2023) è stato inserito un quesito per indagare lo stato attuale dell'interoperabilità dei sistemi informativi nella PA. Il quesito richiedeva se la PA detenesse "banche dati già rese interoperabili o che potrebbero essere rese interoperabili in quanto utili per lo svolgimento di compiti istituzionali di altre pubbliche amministrazioni" (*ivi*, quesito 6.8 del questionario per le unità istituzionali). Su 11.957

2 L'art. 50 del CAD è stato aggiornato diverse volte. Si sottolineano in particolare: l'art. 264, co. 2, lett. b), della L. n.77/2020; l'art. 33, co. 1, lett. a) e l'articolo 34, co. 1, della L. n. 120/2020; l'art. 39, co. 2, lett. a) e b), della L. n. 108/2021; l'art. 18, co. 2 del D.L. n. 13/2023).

istituzioni rispondenti, solamente il 15,8% detiene banche dati rese interoperabili, mentre per il 24,7% le proprie banche dati potrebbero essere rese tali (Bellitti e Fedeli 2023).

In Italia, il tema dello scambio documentale tra diverse amministrazioni è stato trattato nella legge sul procedimento amministrativo, stabilendo che "i documenti attestanti atti, fatti, qualità e stati soggettivi, necessari per l'istruttoria del procedimento, sono acquisiti d'ufficio quando sono in possesso dell'amministrazione procedente, ovvero sono detenuti, istituzionalmente, da altre pubbliche amministrazioni" (art. 18, co. 2, L. 7 agosto 1990, n. 241).

Tale disposizione, rimasta sostanzialmente a lungo inapplicata, è in realtà una norma fondamentale, finalizzata a dare attuazione al principio di 'decertificazione' nei rapporti tra amministrazioni pubbliche e privati; solo con il DPR 445/2000, Testo Unico in materia di documentazione amministrativa, è divenuta la regola generale, declassando a residuale l'utilizzo delle certificazioni, a favore del più agevole sistema delle dichiarazioni sostitutive rese dall'interessato. Questa innovazione nel procedimento amministrativo ha aperto la strada all'uso digitale del documento che, seppure inizialmente ancora pensato come analogico, ha ormai acquisito in sé il concetto di interoperabilità come condizione fondamentale per far transitare in modo automatico e sicuro le informazioni tra PPAA e tra PA e soggetti interessati, anche cittadini. Più di recente, anche la Commissione europea si è espressa sulla questione dell'interoperabilità.

> *I servizi pubblici nazionali dovrebbero essere connessi e andare oltre i confini nazionali per interconnettersi con altri servizi simili a livello UE, contribuendo così al mercato unico digitale. Tutto ciò presuppone l'esistenza di un approccio coordinato a tutti i livelli quando la legislazione viene elaborata, quando le pubbliche amministrazioni organizzano i propri processi lavorativi, quando si gestiscono le informazioni e si sviluppano sistemi informatici per attuare i servizi pubblici. Senza un tale approccio la frammentazione digitale esistente si intensificherà, pregiudicando l'offerta di servizi pubblici connessi in tutta l'UE* (Commissione europea 2017, 2).

La soluzione a questa necessità informativa, volta a favorire la cooperazione tra le PPAA e a semplificare l'erogazione dei servizi, è l'interoperabilità[3] a "tutti i livelli che hanno un impatto sull'erogazione dei servizi pubblici digitali nell'Unione europea" (Commissione europea 2017, 2). A tal proposito, il Quadro europeo di interoperabilità – *European Interoperability Framework* distingue quattro livelli: giuridico, organizzativo, semantico e tecnico.

I quattro aspetti dell'interoperabilità: punti di forza e criticità

L'*interoperabilità giuridica* riguarda la capacità delle PPAA dei diversi Stati membri di collaborare efficacemente nonostante operino in contesti normativi differenti. Essa implica che le normative nazionali non ostacolino la creazione e l'erogazione di servizi pubblici e che siano definiti accordi chiari per affrontare le diversità legislative, anche attraverso l'eventuale introduzione di nuove leggi. Per favorire l'interoperabilità giuridica, il primo passo è effettuare verifiche preventive, cioè analizzare le leggi esistenti per identificare ostacoli come restrizioni sull'uso o la conservazione dei dati, modelli di licenza non uniformi, obblighi tecnologici troppo rigidi, requisiti contraddittori per processi simili, o normative sulla sicurezza ormai obsolete. La coerenza normativa deve essere valutata sia prima dell'adozione sia periodicamente dopo l'entrata in vigore delle norme. Considerando che i servizi pubblici europei si basano molto sui canali digitali, è fondamentale integrare precocemente l'aspetto tecnologico nel processo legislativo tramite una 'verifica digitale'. Inoltre, è necessario assicurare che le informazioni scambiate tra Stati membri mantengano il loro valore giuridico e che siano rispettate le normative sulla protezione dei dati sia nei Paesi di origine sia in quelli di destinazione, anche attraverso accordi aggiuntivi per superare eventuali differenze nell'applicazione delle leggi (Commissione europea 2017, Allegato 2, 25-26).

L'*interoperabilità organizzativa* riguarda il modo in cui le PPAA coordinano e armonizzano i propri processi aziendali, le responsabilità e le aspettative per raggiungere obiettivi comuni e vantaggiosi per tutte le parti coinvolte. Questo vuol dire documentare, integrare o allineare i processi di lavoro e le informazioni scambiate, assicurando così un funzionamento congiunto ed efficace. Tale allineamento richiede una documentazione chiara e concordata dei processi, utilizzando tecniche di modellizzazione condivise. È fondamentale includere le informazioni scambiate in modo che tutte le amministrazioni coinvolte possano comprendere

3 Nel contesto nazionale, il concetto di interoperabilità era già stato definito nel Codice dell'amministrazione digitale come la "caratteristica di un sistema informativo, le cui interfacce sono pubbliche e aperte, di interagire in maniera automatica con altri sistemi informativi per lo scambio di informazioni e l'erogazione di servizi" (art. 1, co. 1, lett. dd).

il processo complessivo, dall'inizio alla fine, e il proprio ruolo in esso (*ivi*, 27-28).

L'*interoperabilità semantica* garantisce che dati e informazioni siano compresi correttamente da tutte le parti coinvolte. Questo aspetto opera su due fronti: sul piano semantico, preserva il significato degli elementi e le loro relazioni, grazie a ontologie, vocabolari controllati e schemi di dati comuni; sul piano sintattico, assicura che formato e struttura delle informazioni siano leggibili e interpretabili da sistemi e soggetti diversi (si veda par. 3). Considerare i dati come un bene pubblico prezioso implica generarli, proteggerli adeguatamente, gestirli, e condividerli, con strumenti che ne standardizzano i contenuti anche attraverso tassonomie, thesauri, *codelists* e strutture riutilizzabili (*ivi*, 28-29).

L'*interoperabilità tecnica* riguarda le applicazioni e le infrastrutture che collegano sistemi e servizi (si veda par. 4). Gli aspetti di questa interoperabilità comprendono soluzioni tecniche che svolgono funzioni di intermediazione tra le diverse applicazioni o di pubblicazione dei dati in formati standard come possono essere ad esempio le API (*Application Programming Interface*). Di fatti, le API sono interfacce software che espongono i propri schemi di funzionamento, consentono di superare le differenze, abilitando l'interoperabilità tra due sistemi informativi, i servizi di interconnessione, i servizi di integrazione dei dati, la presentazione e lo scambio dei dati e i protocolli di comunicazione sicuri (*ivi*, pp. 29-30). Un esempio di soluzione di interoperabilità tecnica è anche la pubblicazione di dati in formato LOD (Linked Open Data), che consentono di esprimere qualsiasi concetto attraverso un modello informativo a triple, ossia 'soggetto-predicato-oggetto'. Tale modello richiama una modalità di rappresentazione della realtà affine a quella della mente umana e permette di collegare risorse provenienti da fonti diverse, favorendo il massimo livello di interoperabilità, la generazione di nuova conoscenza e l'accrescimento del valore informativo dei dati originari.

A tal proposito, Tim Berners-Lee (2010)[4], l'inventore del World Wide Web, propone una classificazione a 5 stelle degli Open Data che rappresenta un modello progressivo per la pubblicazione dei dati, in cui ogni livello aggiunge valore rispetto al precedente e identifica nei LOD il livello superiore (Dietrich 2012):

1. *OL – Open License*: dati pubblicati con licenza aperta in qualsiasi formato. Es.: PDF
2. *RE – Readable and Reusable*: dati in formato strutturato e leggibili da macchine. Es.: XML
3. *OF – Open Format*: dati pubblicati in formati aperti non proprietari. Es.: CSV
4. *URI – Uniform Resource Identifier*: dati identificabili con indirizzi univoci. Es.: Ateco 2025, https://w3id.org/italia/stat/controlled-vocabulary/economy/ateco-2025
5. *LOD – Linked Open Data*: dati collegati ad altri database. Es.: link a fonti di dati nei file RDF.

Un grave ostacolo all'interoperabilità deriva dai sistemi preesistenti. Storicamente, le applicazioni e i sistemi informativi attivi nelle pubbliche amministrazioni sono stati sviluppati in modo bottom-up, per risolvere problemi locali e specifici di dominio (Istat 2021). Ne è risultata una frammentazione delle 'isole ICT'[5], difficili da interoperare. Il Censimento delle istituzioni pubbliche (Istat 2023) ha rilevato oltre 5.000 banche dati, spesso sistemi informativi chiusi, nati per soddisfare esigenze informative all'interno di una specifica cornice istituzionale, ma che non consentono di elaborare e interpretare automaticamente i dati; quindi, di fatto, non sono in grado di abilitare l'interoperabilità. La 'logica a silos' non valorizza i dati e ne impedisce l'analisi che è di supporto ai policy maker per il miglioramento continuo delle politiche pubbliche (si veda par. 5). All'origine di questa frammentazione vi è la mancanza sia di armonizzazione semantica, per via del fatto, ad esempio, che il linguaggio multisettoriale della PA può contenere polisemie e sinonimie che possono ostacolare la comprensione del significato di un dato, ma anche sintattica, senza cioè l'adesione a formati standard univoci condivisi dei dati come TTL, RDF, J-SON (si veda par. 3).

Inevitabili sono le conseguenze soprattutto in termini di coerenza, bassa efficienza operativa e barriere all'innovazione. La mancanza di interoperabilità e di controlli centralizzati di sicurezza facilita la compromissione dei dati: sistemi disconnessi sono più esposti a rischi per la sicurezza. La frammentazione impedisce lo scambio rapido e fluido di dati fra enti, rallentando i processi amministrativi e aumentando i costi gestionali e di manutenzione. L'assenza di un sistema interoperabile, inoltre, limita l'uso integrato di tecnologie avanzate come l'intelligenza artificiale, che necessitano di dati integrati e di qualità per

4 Cfr. Berners-Lee T. (2010), Linked Data is Like a Bag of Chips, in Open, Linked Data for a Global Community, Talk at GovCon 2.0 Expo, May 2010, https://www.w3.org/DesignIssues/BagOfChips.html.

5 La locuzione 'isola ICT' (dove ICT sta per Information and Communication Technology) viene usata in ambito organizzativo e amministrativo per descrivere una situazione in cui i sistemi informativi o tecnologici di un ente o di una struttura operano in modo isolato rispetto ad altri.

funzionare correttamente, con possibili rischi di uso non proporzionato o improprio dei dati. Tutto ciò compromette, evidentemente, la qualità delle analisi e delle decisioni politiche basate su evidenze, con effetti negativi su servizi e politiche pubbliche.

Interoperabilità e governance dei dati: ruoli delle istituzioni e delle piattaforme nazionali ed europee

A partire dagli anni Duemila sono state molteplici le iniziative europee che, in un clima collaborativo tra gli Stati membri, mirano a definire e attuare politiche e linee guida per realizzare servizi digitali interoperabili. Nel 2010, con la comunicazione *Verso l'interoperabilità dei servizi pubblici europei*, la Commissione europea (2010) ha definito una vera e propria *Strategia per l'interoperabilità europea* (SEI), che insieme al *Quadro europeo di interoperabilità* (QEI) è diventata un riferimento chiave per le strategie nazionali. Nel 2017 l'Unione ha adottato lo *European Interoperability Framework* (EIF), con 47 raccomandazioni per migliorare la *governance*, creare relazioni tra organizzazioni, semplificare i processi di supporto ai servizi pubblici e realizzare servizi digitali interoperabili (Commissione europea 2017).

Negli anni successivi le iniziative europee si sono intensificate, finanziando progetti in supercalcolo, intelligenza artificiale, cybersicurezza, competenze digitali. L'Italia si è inserita in questo quadro recependo le direttive e i regolamenti dell'UE. Il *Piano nazionale di ripresa e resilienza* (PNRR) presenta tra gli obiettivi quello del raggiungimento della completa interoperabilità dei dataset e dei servizi chiave tra le PPAA centrali e locali, al fine di valorizzare il loro capitale informativo (PNRR, sub-investimento M1C1_1.3.1, Piattaforma nazionale digitale dei dati), mentre l'Agenzia per l'Italia Digitale ha sviluppato il Modello di interoperabilità (ModI), basato sulle sue Linee guida (determinazione AgID 11 ottobre 2021 n. 547) in coerenza con i principi dello EIF, come parte del Piano triennale per l'informatica nella Pubblica amministrazione (AgID 2023a).

Il Piano triennale per l'informatica nella Pubblica amministrazione 2024-2026 (aggiornato a gennaio 2025) è un documento di programmazione strategica che contiene sia indirizzi generali sia indicazioni operative sul come e quando gli enti possano integrare il loro sistema informativo con i servizi, le piattaforme e le infrastrutture erogate a livello centralizzato. L'obiettivo è garantire servizi pubblici digitali integrati, infrastrutture digitali sicure e sostenibili, semplificando e dando valore aggiunto all'esperienza dell'utente. Tra i principi guida del Piano è richiamato il principio AgID dell'interoperabilità *by design* e *by default* (API-*first*) secondo cui "i servizi pubblici devono essere progettati in modo da funzionare in modalità integrata e attraverso processi digitali collettivi" (AgID 2023a, 12).

Nel suo ultimo aggiornamento (AgID 2026) si consolida il ruolo centrale del Responsabile per la transizione digitale (RTD) e del relativo ufficio (UTD), istituito in forma singola o associata, in ottemperanza all'art. 17 del CAD, sottolineandone l'ulteriore funzione di promotori di attività di collaborazione, condivisione e valorizzazione di conoscenze e buone pratiche tra RTD di enti diversi. Tali azioni sono finalizzate a fare della PA un "ecosistema digitale amministrativo" (AgID 2026, 25) con dotazione di risorse umane, strumentali e finanziarie volte a ridefinire l'assetto organizzativo e operativo degli uffici, improntato sul dettato costituzionale dell'art. 97, co. 2 (dal punto di vista di imparzialità e buon andamento dell'azione amministrativa), a lungo considerato un "irrilevante giuridico" (De Seta 1985, 48), nonché all'art. 41 della Carta dei diritti fondamentali dell'Unione europea (c.d. Carta di Nizza) relativo al diritto ad una buona amministrazione, inteso come pretesa di imparzialità procedimentale e processuale nei confronti di ogni individuo. La consultazione pubblica *Verso il nuovo Codice dell'Amministrazione digitale (CAD)*, avviata a maggio 2026 dal Dipartimento per le Riforme istituzionali, mira a raccogliere contributi anche in vista dell'evoluzione della disciplina tesa alla realizzazione di e-service in grado di implementare pienamente il principio *once only* "poiché abilita il potere istruttorio di acquisizione d'ufficio delle amministrazioni pubbliche ogni qualvolta occorrano dati e informazioni che siano già in possesso dell'amministrazione pubblica procedente o di altre per lo svolgimento del procedimento amministrativo (L. 241/1990, art. 18)" (Alberti 2022, 477).

La Piattaforma digitale nazionale dati è stata istituita dall'articolo 50-ter del Codice dell'amministrazione digitale e pubblica il Catalogo delle API che espone come *e-service* per l'interscambio di dati tra PPAA. Gli e-service rappresentano un servizio digitale, spesso basato su API pubblicate dagli enti erogatori, che consente lo scambio autorizzato, automatizzato e sicuro di dati tra ente erogatore ed ente fruitore. I dati ottenuti attraverso questo tipo di accesso sono annotati semanticamente tanto da renderli inequivocamente integrabili. Strettamente connesso alla PDND, il *National Data Catalog* (si veda par. 4), frutto di un accordo tra il Dipartimento per la Trasformazione digitale (DTD), in qualità di soggetto titolare e l'Istat,

in qualità di soggetto attuatore[6], è la piattaforma che pubblica risorse semantiche *machine readable* (ontologie, vocabolari controllati e schemi di dati) attraverso le quali è possibile eseguire l'annotazione semantica. La realizzazione del Catalogo richiede la collaborazione delle pubbliche amministrazioni, in quanto titolari delle competenze specialistiche relative ai rispettivi domini da modellare semanticamente. Le amministrazioni sono inoltre responsabili della validità dei propri asset semantici, sia quando operano autonomamente sia quando si avvalgono di diverse forme di *stewardship* con l'Istat.

3. Strumenti per l'interoperabilità semantica

Se l'interoperabilità tecnica costituisce la condizione necessaria per lo scambio di informazioni tra sistemi informativi, quella semantica rappresenta il livello che consente ai dati di essere effettivamente comprensibili, coerenti e riutilizzabili al di là del contesto in cui sono stati prodotti (Euzenat e Shvaiko 2013). In assenza di strumenti adeguati, l'integrazione dei dati risulta complessa: terminologie differenti, schemi concettuali incompatibili e ambiguità linguistiche rendono difficile la condivisione e l'analisi congiunta delle informazioni (Gruber 1995; Guarino 1998).

Il *National Data Catalog* rappresenta la concretizzazione della dimensione semantica della Piattaforma digitale nazionale dati (Aracri *et al.* 2024). Il portale del Catalogo, *schema.gov.it* (si veda paragrafo successivo), raccoglie e organizza le risorse semantiche della PA, classificandole in tre categorie principali:

1. le *ontologie*, che forniscono una rappresentazione strutturata dei principali domini informativi della PA, includendo persone, luoghi, organizzazioni, servizi e beni culturali ecc., e permettono di definire le relazioni tra questi concetti in modo coerente e condiviso;
2. i *vocabolari controllati*, che garantiscono uniformità terminologica, riducono l'ambiguità dei termini e facilitano la comunicazione tra sistemi e operatori diversi, creando un linguaggio comune per descrivere dati e processi amministrativi;
3. gli *schemi di dati*, che descrivono in modo chiaro e standardizzato la struttura dei *dataset*, specificando campi, tipi di dati, relazioni e vincoli, e fungono da vero e proprio 'contratto' tra chi produce e chi utilizza le informazioni, assicurando interoperabilità e coerenza nell'uso dei dati pubblici digitali[7].

Ogni risorsa è pubblicata con identificatori persistenti (URI) e formati e linguaggi standard (OWL, RDF, JSON-LD, YAML), così da permettere la piena interoperabilità tra amministrazioni e la lettura automatizzata da parte dei sistemi informativi.

Ontologie: rappresentazione formale della conoscenza

Un'ontologia è una rappresentazione esplicita e condivisa di un dominio di conoscenza, composta da concetti, proprietà e relazioni logiche (Studer *et al.* 1998). Il suo scopo è offrire un quadro coerente e interpretativamente stabile del significato dei dati, permettendo a persone e sistemi di riferirsi agli stessi concetti in modo univoco. Attraverso classi, relazioni e istanze, le ontologie consentono di creare grafi di conoscenza (*knowledge graphs*), in cui i concetti e le loro connessioni sono rappresentati come nodi e archi. Le ontologie permettono di modellare relazioni complesse e multidimensionali, favorendo la condivisione della conoscenza in domini diversi. Esse costituiscono la base concettuale del *Web Semantico* e dei *Linked Data* (Berners-Lee *et al*. 2001), permettendo l'interconnessione tra basi di dati distribuite e la gestione intelligente dell'informazione. Le ontologie facilitano una comprensione condivisa del dominio e rendono esplicite le assunzioni alla base dei modelli informativi. Ciò migliora la qualità dei metadati, la tracciabilità e la possibilità di analisi avanzata. Inoltre, la loro struttura flessibile consente di adattare e ampliare i modelli nel tempo senza comprometterne la coerenza, garantendone continuità e riutilizzo.

Il *Web Ontology Language* (OWL) è lo standard del W3C[8] per esprimere ontologie sul Web. Basato su logiche descrittive, OWL consente

6 Art. 5, co. 6, D.Lgs. n. 50/2016, per la realizzazione della misura 1.3.1 Catalogo Nazionale Dati della Missione M1 – Componente C.1 – Asse 1 del PNRR. Il PNRR, approvato il 13 luglio 2021 con Decisione di esecuzione del Consiglio (CID) contenente in un allegato gli obiettivi e i traguardi per ciascuna missione, richiama alla Missione 1 – componente 1 Digitalizzazione della Pubblica amministrazione le misure volte a migliorare la digitalizzazione della PA, che si articolano in sette investimenti e tre riforme. Gli investimenti sono diretti in particolare a: 1) razionalizzare e consolidare le infrastrutture digitali esistenti della PA; 2) promuovere la diffusione del cloud computing; 3) realizzare il principio *once only* e l'accessibilità dei dati tramite un catalogo centrale di 'connettori automatici' (cosiddette API - Application Programming Interface), con particolare attenzione all'armonizzazione e all'interoperabilità delle piattaforme e dei servizi di dati.

7 Al momento dell'ultima revisione di questo articolo, agosto 2026, il Catalogo conta 414 risorse semantiche: 81 ontologie, 206 vocabolari controllati e 127 schemi di dati.

8 Il W3C, acronimo di *World Wide Web Consortium*, è un'organizzazione internazionale fondata da Tim Berners-Lee nel 1994, responsabile dello sviluppo degli standard aperti per il web semantico. Tra i suoi principali standard ci sono RDF, OWL e SKOS, utilizzati per rappresentare concetti, relazioni e vocabolari controllati in modo formalizzato e leggibile sia da persone sia da macchine.

di definire classi, proprietà e relazioni in modo computabile, assicurando coerenza semantica e possibilità di inferenza automatica. I *reasoner* OWL consentono di verificare la consistenza dei modelli e di dedurre nuove informazioni dai dati, sostenendo l'integrazione semantica in contesti eterogenei come, ad esempio, amministrazione pubblica, ricerca o sanità. Le ontologie incorporano relazioni semantiche tra concetti, permettendo il ragionamento automatico sui dati, soprattutto nei database semantici a grafo (*RDF triplestore*) che le usano come schemi di riferimento. Funzionano come un cervello, elaborando e collegando concetti in modo simile alla percezione umana, e favoriscono una navigazione coerente tra le informazioni. Sono facilmente estendibili, consentendo l'aggiunta di nuove relazioni senza compromettere l'integrità del modello, ed evolvono con la crescita dei dati senza impatti negativi sui sistemi. Possono rappresentare dati di diversa natura (strutturati o meno), facilitando integrazione, analisi concettuale, *text mining* e *data-driven analytics*. Definiscono i termini di un dominio, migliorando la gestione della conoscenza e la rappresentazione dei rapporti tra concetti.

L'approccio *Ontology-Based Data Management* (OBDM) (Lenzerini 2011; Calvanese *et al.* 2011) consente di integrare dati eterogenei utilizzando un'ontologia come schema concettuale globale. Il modello prevede tre livelli: l'*ontology layer* rappresenta la conoscenza di dominio; il *mapping layer* collega i concetti dell'ontologia alle sorgenti fisiche; il *data source layer* contiene i dati effettivi. Le interrogazioni formulate in linguaggio SPARQL[9] vengono tradotte in *query* SQL[10] sulle fonti sottostanti, consentendo accesso e integrazione in tempo reale. Questo approccio separa nettamente il livello semantico da quello fisico, facilitando la gestione evolutiva dei sistemi informativi e l'adattamento a nuove normative e contesti organizzativi (Aracri *et al.* 2024).

Vocabolari controllati: infrastruttura semantica per l'interoperabilità dei dati

I vocabolari controllati sono insiemi organizzati di termini o concetti che garantiscono coerenza semantica e uniformità linguistica nella descrizione dei dati (Harpring 2010). Essi specificano le relazioni tra i termini – gerarchiche, associative o equivalenti – riducendo ambiguità e variabilità terminologica. In un ambiente digitale complesso, i vocabolari controllati agiscono come 'ponti semantici' tra sistemi e domini diversi. Nelle *digital humanities*, ad esempio, rappresentano strumenti di mediazione tra linguaggio specialistico e rappresentazione formale, permettendo di condividere conoscenza culturale e disciplinare in formati leggibili da macchina. Nel settore pubblico, i vocabolari controllati favoriscono la standardizzazione di denominazioni, classificazioni e metadati. Ad esempio, vocabolari comuni per 'tipologia di ente', 'tipo di servizio' o 'categoria di documento' permettono di integrare informazioni provenienti da amministrazioni diverse, migliorando la qualità e la fruibilità dei dataset.

Con il web semantico, i vocabolari controllati hanno acquisito un ruolo computazionale: le relazioni semantiche esplicite possono essere interpretate automaticamente dai sistemi, rendendo possibile la classificazione e la ricerca basata sul significato. A tale scopo il W3C (2009; 2024) ha definito lo standard SKOS (*Simple Knowledge Organization System*) che rappresenta concetti (skos:Concept) con etichette linguistiche, relazioni semantiche e identificatori persistenti. SKOS costituisce un modello intermedio tra glossari informali e ontologie complesse: mantiene la semplicità descrittiva dei thesauri, ma offre compatibilità con i linguaggi semantici del Web (ISO 2011). In ambito amministrativo, SKOS consente di allineare vocabolari tra enti e favorire la costruzione di cataloghi nazionali condivisi. È inoltre compatibile con RDF e OWL, permettendo l'integrazione con ontologie e *knowledge graph*. In questo modo, i vocabolari controllati non solo garantiscono coerenza linguistica, ma diventano componenti attivi delle infrastrutture di interoperabilità semantica, facilitando il riuso dei dati e la convergenza semantica tra domini diversi (Zaytseva e Ďurčo 2020).

Schemi di dati: struttura, armonizzazione e contratti semantici

Accanto a ontologie e vocabolari, gli schemi di dati definiscono la struttura logica dei dataset scambiati ad esempio attraverso una API: specificano campi, tipi di dato, relazioni e vincoli. Essi traducono la conoscenza condivisa in modelli implementabili, fungendo da 'contratto semantico' tra chi produce e chi utilizza i dati (Hendler *et al.* 2020). Gli schemi permettono di validare e documentare i dataset,

9 SPARQL (SPARQL Protocol and RDF Query Language) è un linguaggio di interrogazione progettato per estrarre e manipolare dati memorizzati in formato RDF (Resource Description Framework). Viene utilizzato nel contesto del web semantico e delle ontologie, consentendo di interrogare grafi di conoscenza e basi di dati semantiche, come quelle che descrivono relazioni tra risorse e concetti.

10 SQL (Structured Query Language) è un linguaggio standard per la gestione e l'interrogazione di database relazionali. Permette di creare, modificare, leggere ed eliminare dati organizzati in tabelle con relazioni predefinite.

assicurando coerenza interna e interoperabilità tra sistemi. Nella PA italiana, RDF Schema, OWL, JSON-LD, XML Schema sono esempi di modelli comuni che rendono compatibili i dati di enti diversi (AgID 2023b). Grazie a questi standard è possibile descrivere in modo uniforme entità territoriali, organizzazioni pubbliche e servizi digitali, facilitando il riuso e l'analisi dei dati. Negli ultimi anni si sono diffusi anche strumenti come JSON Schema e SHACL (*Shapes Constraint Language*), che consentono di verificare automaticamente la conformità dei dati alle specifiche semantiche. Tali linguaggi permettono di esprimere regole di validazione e vincoli semantici direttamente nei dataset, rafforzando la qualità e l'affidabilità dei dati scambiati.

In Europa, gli schemi di dati sono parte integrante delle politiche di interoperabilità promosse dallo *European Data Strategy* e dal *Data Governance Act*, che incoraggiano l'uso di modelli aperti e riusabili (si veda par. 2). In questo contesto, lo standard DCAT (*Data Catalog Vocabulary*) – adottato anche in Italia – consente di descrivere dataset e cataloghi in modo omogeneo, migliorando la reperibilità e la compatibilità semantica. Gli schemi di dati possono essere visti come l'anello di congiunzione tra la dimensione semantica (ontologie, vocabolari) e quella tecnica (formati e protocolli di scambio). Essi sostengono la costruzione di ecosistemi informativi interoperabili, nei quali i dati diventano realmente condivisibili e comprensibili da persone e macchine. Inoltre, definiscono un linguaggio comune che semplifica la cooperazione tra amministrazioni e imprese, contribuendo alla realizzazione di uno spazio europeo dei dati basato sui principi FAIR (Wilkinson *et al.* 2016; Commissione europea 2020).

4. Il caso di studio: il Catalogo nazionale per la semantica dei dati

Il tema dell'interoperabilità dei dati pubblici ha radici lontane e si lega alle esperienze condotte dagli Istituti di statistica europei già vent'anni fa. Il problema non era soltanto quello di condividere dati, ma di garantirne qualità, stabilità e soprattutto comprensione reciproca. Fu introdotto allora un approccio, basato su vocabolari condivisi e risorse semantiche, che consentì di superare differenze linguistiche e strutturali tra i vari Paesi, dando vita ad un ecosistema statistico integrato. In Italia, nel campo della PA, questa stessa esigenza si è tradotta inizialmente con il progetto OntoPiA, sistema di ontologie e vocabolari controllati promosso da AgID nel 2019. Successivamente, attraverso fondi *NextGenerationEU*, dunque nell'ambito del PNRR, nasce *schema.gov.it*, come parte della PDND. L'obiettivo è realizzare un'infrastruttura che consenta a tutte le PPAA di scambiare dati in modo standardizzato, eliminando ridondanze e incoerenze. La missione di *schema.gov.it* è duplice: da un lato assicurare l'interoperabilità semantica, dall'altro tradurre i principi europei (Commissione europea 2017; 2022; 2024) in strumenti concreti per la PA italiana. L'iniziativa punta a rafforzare la trasformazione digitale della PA, massimizzando il valore informativo dei dati e riducendo i costi di gestione. Il principio *once only*, cardine dell'interoperabilità, permette ai cittadini e alle imprese di fornire i dati soltanto all'amministrazione competente che si prenderà carico di renderla disponibile a tutte le altre PPAA che sono autorizzate a gestire quei dati.

Il *Catalogo nazionale per la semantica dei dati* è in linea con i principi FAIR (Wilkinson *et al.* 2016) e con le raccomandazioni sul riuso dei dati pubblici presenti nella *European strategy for data*, lanciata dalla Commissione europea nel 2020. Questo richiede un lavoro di standardizzazione semantica, ossia di armonizzazione dei concetti, delle definizioni e delle regole di scambio. Il Catalogo risponde a questa sfida attraverso la piattaforma *schema.gov.it*, raccogliendo schemi dati, vocabolari controllati e ontologie che consentono di descrivere e utilizzare in modo coerente informazioni su persone, organizzazioni, luoghi e servizi. Le risorse raccolte in NDC sono organizzate per strumenti semantici, categorie e titolari della risorsa[11].

Il progetto si fonda su collaborazione, riuso e trasparenza: ogni ente è chiamato a contribuire con il proprio *know-how*, condividendo conoscenze che diventano patrimonio comune. Il portale consente di effettuare ricerche per parola chiave, dominio e tipo di risorsa e mette a disposizione strumenti di supporto come validatori di metadati ed *editor* semantici. Il processo di raccolta è basato su *harvesting* automatico dai *repository Git* delle amministrazioni, che assicura aggiornamenti continui e riduce il rischio di errori manuali (Bellitti e Fedeli 2023). Ogni risorsa è identificata da un URI (*Uniform Resource Identifier*) persistente, una URL legata al servizio *w3id.org*, che ne garantisce stabilità, rintracciabilità e unicità.

L'Istat svolge il ruolo di soggetto attuatore del progetto, centrale nella definizione di metodologie,

11 Esempi di risorse: Ontologia degli Indirizzi/Luoghi: https://w3id.org/italia/onto/CLV; Vocabolario Controllato: Ateco 2025 - Classificazione delle attività economiche: https://w3id.org/italia/stat/controlled-vocabulary/economy/ateco-2025; Consultazione Persona Fisica API: https://raw.githubusercontent.com/InailUfficio5/inail-ndc/main/assets/schemas/persona-fisica/latest/persona-fisica.oas3.yaml.

supporto tecnico e garanzia della qualità. La metodologia Istat per il Catalogo si fonda su due pilastri coordinati: l'analisi concettuale, dove esperti giuridici estraggono dalle norme definizioni univoche per garantire la legittimità dello scambio, e la modellazione tecnica, che traduce tale conoscenza in ontologie, vocabolari e schemi mediante linguaggi formali (OWL, SKOS) e identificatori persistenti (URI/IRI). Il processo segue un modello di gestione incrementale che assicura la coerenza semantica tra i sistemi informativi attraverso l'integrazione di concetti *core* e di dominio (Bellitti e Fedeli 2024; 2025). L'Istituto collabora con il Dipartimento per la trasformazione digitale, che è il soggetto titolare e coordinatore del progetto, e con AgID e CNR, soggetti coinvolti sul tema dell'interoperabilità. Ogni ente contributore è responsabile della gestione e dell'aggiornamento delle proprie risorse semantiche, nel rispetto di un quadro comune di regole, in un contesto di governance diffusa.

Sono previste tre modalità di contribuzione alle risorse semantiche: la *semantic stewardship* dell'Istat, che affianca gli enti privi di competenze interne nell'analisi concettuale e nella modellazione ontologica; lo *sviluppo autonomo delle risorse*, con supervisione e validazione dei metadati; la *formazione da parte dell'Istat per sviluppo schemi dati*, che consente di acquisire competenze tecniche e metodologiche attraverso l'uso dello Schema Editor e di altri strumenti (Bellitti e Fedeli 2025). Questa architettura collaborativa consente di valorizzare sia le grandi amministrazioni, come INPS (Istituto nazionale previdenza sociale) e Ministero dell'Interno, che godono di maggior autonomia, sia enti più piccoli, creando una rete distribuita di conoscenze e pratiche comuni.

Il Catalogo adotta misure stringenti per assicurare qualità, sicurezza e coerenza. La qualità delle risorse è garantita da processi di *auditing* manuale e da validazioni automatiche durante l'*harvesting*. Sul piano della sicurezza, la piattaforma rispetta lo standard ISO/IEC 27002 (ISO/IEC 2022) per integrità e riservatezza e applica le best practice OWASP (*Open Web Application Security Project)*. Inoltre, il software NDC viene sottoposto periodicamente a ulteriori controlli di sicurezza mediante strumenti dedicati (Polizzi e Letardi 2023).

Tra gli strumenti tecnici disponibili nel Catalogo vi sono: il *Validatore Metadati*, che controlla la corretta descrizione delle risorse; lo *Schema Editor*, che consente la creazione e l'annotazione semantica degli schemi dati; lo *SPARQL endpoint*, che permette interrogazioni avanzate.

Diversi esempi mostrano l'impatto concreto dell'iniziativa. Il Ministero dell'Interno, insieme all'ANPR (Anagrafe nazionale della popolazione residente), ha definito una modellazione condivisa di 'persona residente', adottata da vari enti e utile a evitare ridondanze informative. La Camera e il Senato stanno armonizzando le proprie ontologie in una semantica parlamentare unificata, basata sulla Costituzione e su vari regolamenti, che migliora l'accessibilità e la coerenza dei dati legislativi. L'INPS ha avuto un ruolo pionieristico, pubblicando schemi dati annotati semanticamente, e sperimentando strumenti di annotazione e validazione. L'Agenzia del Demanio ha richiesto la *stewardship* completa dell'Istat, ottenendo supporto nella modellazione e formazione del proprio personale, con la creazione di nuove risorse semantiche di dominio. Questi casi dimostrano come il Catalogo consenta sia standardizzazione sia innovazione, riducendo la complessità e aumentando l'efficienza dei processi.

Schema.gov.it offre valore concreto a più categorie di utenti. Le pubbliche amministrazioni ottengono uniformità, integrazione tra sistemi e attuazione del principio *once only*. Gli sviluppatori possono riutilizzare risorse già disponibili per creare API interoperabili e servizi digitali coerenti; il Catalogo mette a loro servizio una serie di ontologie *core*, risorse semantiche relative a concetti di carattere generale e trasversale, come l'ontologia della Persona (*Core Person Vocabulary* - CPV) sviluppata dall'Istat, che è alla base del riutilizzo, dell'interoperabilità e della creazione di *Linked Open Data* (LOD) e che mette in connessione più domini (Puglisi e Radini 2025). I ricercatori beneficiano di un patrimonio semantico ampio e certificato per studi e analisi. Infine, i fornitori di soluzioni ICT e i tecnici trovano strumenti e standard per gestire il ciclo di vita dei servizi, riducendo costi e tempi di sviluppo. Tutti questi attori contribuiscono a creare un ecosistema più efficiente, trasparente e orientato alla realizzazione di servizi standardizzati per il cittadino.

Un aspetto innovativo riguarda l'uso dei dati semanticamente arricchiti per alimentare sistemi di intelligenza artificiale. Il Catalogo fa parte della quinta sezione della Strategia italiana per l'intelligenza artificiale 2024-2026 e rappresenta un importante strumento per promuovere l'applicazione dell'IA all'interoperabilità, un ambito emergente identificato dalla Commissione europea con il termine AI4IOP (Davide 2024). L'IA permette di migliorare l'interoperabilità tra enti, superando frammentazioni e incompatibilità. Attualmente, l'Istat sta sperimentando modelli basati su *AI summarisation chain*, per sintetizzare dati, e su *AI data analytics*, per individuare anomalie e produrre analisi predittive. L'adozione di un approccio *human in the loop*, tipico del settore pubblico, garantisce controllo umano sulle decisioni e riduce i rischi

legati a errori o *bias* dei modelli (Bellitti e Fedeli 2024). In prospettiva, l'integrazione tra *Open Data*, API, ontologie e IA consentirà la costruzione di *knowledge graph* e sistemi di conoscenza più intelligenti, capaci di supportare processi decisionali affidabili e trasparenti, attraverso prodotti avanzati di IA come il *reasoner* automatico (Lenzerini 2023).

5. Impatti attesi sull'uso dei dati amministrativi per le politiche pubbliche

L'interoperabilità e la standardizzazione semantica dei dati amministrativi rappresentano condizioni abilitanti fondamentali per trasformare tali risorse in strumenti operativi al servizio dei processi di policy making, di monitoraggio e di valutazione delle politiche pubbliche (Stiglitz 1999). La letteratura sottolinea da tempo come un ecosistema informativo ben orchestrato sia in grado di aumentare la capacità delle istituzioni di prendere decisioni basate su evidenze solide, favorire l'innovazione e rafforzare la fiducia dei cittadini nei confronti delle istituzioni (Dawes 1996; Ubaldi 2013; OECD 2019). La possibilità di integrare basi di dati diverse, strutturate secondo schemi comuni e arricchite da risorse semantiche, consente di ridurre errori di integrazione e duplicazioni informative, migliorando al contempo la qualità e la coerenza delle informazioni disponibili. Studi recenti mostrano come l'adozione di ontologie e schemi semantici comuni faciliti l'integrazione di dati eterogenei, riduca le inconsistenze e supporti il riuso dei dati in contesti multipli (Bizer *et al.* 2023; Ranatunga *et al.* 2025).

Un primo impatto riguarda la tempestività e la precisione delle politiche pubbliche. L'accesso a dati interoperabili permette di ridurre il tempo che intercorre tra l'osservazione di un fenomeno e la sua traduzione in un intervento normativo o amministrativo, favorendo analisi più frequenti e localizzate e consentendo stime affidabili a livello regionale e sub-regionale (Rivas e Crowley 2018). La disponibilità di informazioni aggiornate e coerenti è particolarmente rilevante in settori come il lavoro, la sanità o il welfare, dove la rapidità di reazione può determinare l'efficacia stessa dell'intervento. Parallelamente, l'adozione di standard condivisi e del principio *once only,* promosso a livello europeo (Commissione europea 2020), comporta un significativo aumento di efficienza amministrativa.

L'interoperabilità semantica dei dati ha implicazioni dirette anche sul fronte della valutazione delle politiche pubbliche. La possibilità di costruire serie storiche coerenti e di effettuare confronti longitudinali, transregionali e transnazionali rende più robuste le valutazioni di impatto e consente di applicare metodologie quasi-sperimentali in contesti reali (OECD 2019; Zuiderwijk *et al.* 2021). Ciò si traduce in evidenze empiriche più solide, indispensabili per orientare decisioni informate e allocazioni più efficaci delle risorse pubbliche. Non meno importante è l'effetto sulla trasparenza e sulla *accountability*: dati ben documentati, accessibili e accompagnati da metadati standardizzati favoriscono audit indipendenti, studi replicabili e processi di responsabilizzazione democratica, contribuendo a rafforzare la fiducia dei cittadini nelle istituzioni (Meijer 2014; Commissione europea 2017).

La disponibilità di schemi condivisi e vocabolari controllati ha inoltre un impatto positivo sull'innovazione e sullo sviluppo di servizi digitali avanzati. La riduzione dei costi di integrazione e la maggiore uniformità semantica dei dati facilitano la creazione di *dashboard* interattive, sistemi predittivi e applicazioni di intelligenza artificiale responsabile (AgID 2023b). Tali strumenti, fondati su dati amministrativi interoperabili, aprono la strada a nuove forme di analisi multidisciplinare che permettono di collegare politiche tradizionalmente separate – come quelle del lavoro, della sanità e dell'assistenza sociale – e di affrontare fenomeni complessi in un'ottica integrata. Questo approccio offre anche la possibilità di monitorare con maggiore precisione le disuguaglianze territoriali e sociali, identificando *gap* tra regioni e gruppi di popolazione e orientando interventi mirati che promuovano politiche più inclusive (Istat 2022).

Gli impatti attesi, tuttavia, non sono automatici né esenti da rischi. La loro realizzazione dipende dalla qualità degli standard adottati, dall'esistenza di una *governance* sostenibile e condivisa, dalla disponibilità di risorse infrastrutturali e competenze adeguate, nonché dal pieno rispetto delle normative sulla protezione dei dati personali, in particolare del Regolamento generale sulla protezione dei dati (GDPR 2016; Commissione europea 2017). Senza questi presupposti, il potenziale dei dati amministrativi rischia di rimanere inespresso o di produrre effetti distorti.

In questa cornice si inserisce il progetto di *schema.gov.it*, di cui al paragrafo precedente. La costruzione di un ecosistema di dati amministrativi interoperabile e semanticamente standardizzato che fornisce gli strumenti per realizzare una PA più tempestiva, efficiente e inclusiva, capace di fornire dati e analisi di fenomeni solidi per l'erogazione di servizi amministrativi e informazioni per valutare le politiche, di abilitare innovazione digitale e di rafforzare la fiducia dei cittadini. Gli studi empirici condotti da organismi internazionali come l'IMF (*International Monetary Fund*) e l'OECD (*Organisation for Economic Co-operation and*

Development) hanno già mostrato come tali benefici si traducano in risparmi economici significativi, in una riduzione dei costi di transazione per cittadini e imprese e in una maggiore qualità delle statistiche ufficiali (Rivas e Crowley 2018; OECD 2019). L'impatto atteso non riguarda dunque soltanto l'efficienza amministrativa, ma investe la dimensione stessa del rapporto tra Stato, cittadini e imprese, rafforzando la democrazia e la capacità del sistema pubblico di affrontare le sfide emergenti.

Conclusioni

L'analisi condotta in questo contributo ha mostrato come l'interoperabilità dei dati amministrativi costituisca una condizione imprescindibile per la piena valorizzazione del patrimonio informativo pubblico e per fornire informazioni integrate da sfruttare per la costruzione di politiche basate su evidenze concrete. In un contesto europeo caratterizzato da una crescente attenzione alla gestione intelligente dei dati, la *European Data Strategy* e i principi FAIR (Wilkinson *et al.* 2016) tracciano una direzione chiara: trasformare i dati in un'infrastruttura abilitante per la conoscenza e l'innovazione. In tale quadro, l'esperienza italiana rappresentata da *schema.gov.it*, unica per ampiezza dei domini analizzati e per i servizi forniti, assume un ruolo strategico, ponendosi come piattaforma di riferimento per la definizione, la pubblicazione e l'allineamento di ontologie, vocabolari controllati e schemi di dati, strumenti indispensabili per garantire coerenza semantica, standardizzazione e riusabilità delle informazioni.

Il lavoro ha evidenziato come la standardizzazione semantica e la governance condivisa di dati e metadati costituiscano elementi centrali non solo sul piano tecnico, ma anche su quello culturale, organizzativo e istituzionale della trasformazione digitale della Pubblica amministrazione. L'evoluzione del CAD mostra infatti il passaggio da una concezione del dato come patrimonio settoriale del singolo ente a una visione ecosistemica del patrimonio informativo pubblico, fondata su interoperabilità, collaborazione e riuso (cfr. sottopar. *Interoperabilità e governance dei dati: ruoli delle istituzioni e delle piattaforme nazionali ed europee*). In tale prospettiva, l'interoperabilità semantica si configura come fattore abilitante dell'efficacia amministrativa, della trasparenza e del diritto a una buona amministrazione, in coerenza con l'art. 97 della Costituzione e con l'art. 41 della Carta dei diritti fondamentali dell'UE, la cui effettività passa oggi anche attraverso la digitalizzazione dei processi pubblici (Galetta 2021). La sperimentazione condotta nello studio *Verso l'interoperabilità. Una sperimentazione di interconnessione tra strumenti e piattaforme ICT dell'Inapp* conferma concretamente come l'interconnessione tra sistemi eterogenei possa generare valore conoscitivo, supportare analisi più avanzate e rafforzare politiche pubbliche basate su evidenze (Inapp 2025). In prospettiva, la sfida sarà tradurre i principi di interoperabilità semantica in obblighi cooperativi più vincolanti, capaci di superare la frammentazione delle 'isole ICT' e di favorire infrastrutture pubbliche integrate, aperte e orientate anche all'uso responsabile di strumenti analitici avanzati e di intelligenza artificiale.

La piena realizzazione di questo scenario incontra sfide rilevanti di natura tecnologica, giuridica, organizzativa e culturale. Da un lato, è necessario consolidare piattaforme scalabili, sicure; dall'altro, occorre garantire una gestione equilibrata tra apertura dei dati e tutela della privacy, in linea con il GDPR e con i più recenti indirizzi europei in materia di data governance. Sul piano organizzativo, la mancanza di un coordinamento strutturato tra le diverse amministrazioni può ostacolare l'adozione di standard comuni: per questo, la creazione di modelli di governance condivisi, che prevedano ruoli chiari e procedure di aggiornamento costanti, rappresenta una priorità imprescindibile. A ciò si aggiunge la dimensione culturale, che richiede un investimento nelle competenze digitali e semantiche sia nella PA sia nella comunità scientifica, per stimolare una più ampia comprensione del valore dei dati come bene pubblico. Sul piano economico-sociale, l'impatto di un quadro regolatorio così delineato è dirompente. L'abbattimento dei costi di transazione e la riduzione dei tempi procedurali per cittadini e imprese si traducono in un incremento diretto di competitività per il sistema-Paese. Inoltre, un ecosistema arricchito semanticamente permette una mappatura tempestiva delle disuguaglianze territoriali, orientando in modo puntuale gli interventi di welfare.

Guardando al futuro, lo sviluppo di *schema.gov.it* come infrastruttura nazionale per l'interoperabilità semantica può costituire la base di un ecosistema integrato di conoscenza, in grado di connettere *linked open data*, *e-service* e IA in una rete coerente e dinamica, aperta a collaborazioni con università e centri di ricerca e allineata con le principali iniziative europee. In questa direzione, la sfida più grande sarà mantenere un equilibrio tra innovazione tecnologica e sostenibilità istituzionale. Ciò implica l'adozione di strumenti di monitoraggio continuo della qualità dei dati, la promozione di spazi di sperimentazione per l'uso sicuro dei dati in ambito IA e l'incentivazione di modelli partecipativi in cui le amministrazioni siano co-progettisti dell'infrastruttura.

In conclusione, l'interoperabilità dei dati non rappresenta un traguardo tecnico, ma un

processo evolutivo e collaborativo essenziale per trasformare il patrimonio informativo pubblico in una leva strategica di trasparenza, equità sociale e partecipazione democratica, in piena linea con le ambizioni della *European Data Strategy* e con gli obiettivi di un'Europa digitale, inclusiva e sostenibile.

Bibliografia

AgID (2026), *Piano Triennale per l'informatica nella Pubblica Amministrazione. Edizione 2024-2026*, *Aggiornamento 2026,* Roma, AgID

AgID (2023a), *Piano Triennale per l'informatica nella Pubblica Amministrazione. Edizione 2024-2026*, Roma, AgID

AgID (2023b), *Linee Guida sull'interoperabilità tecnica delle Pubbliche Amministrazioni,* Roma, AgID

Alberti I. (2022), La creazione di un sistema informativo unitario pubblico con la Piattaforma digitale nazionale dati, *Istituzioni del Federalismo*, XLIII, n.2, pp.473-494

Amico A., D'Alessandro G., Decataldo A. (2018), Advantages of Administrative Data: Three Analyses of Students' Careers in Higher Education, in Lauro N., Amaturo E., Grassia M., Aragona B., Marino M. (eds.), *Data Science and Social Research. Studies in Classification, Data Analysis, and Knowledge Organization*, Cham, Springer, pp.131-139

Aracri R.M., Frisardi D., Radini R., Santarelli V. (2024), Empowering e-services through the Semantic Web, in *CEUR Workshop Proceedings*, Vol.3762, pp.452-457

Bellitti G., Fedeli M. (a cura di) (2025), *Catalogo Nazionale Dati. Metodologia per l'interoperabilità semantica*, Roma, Istat

Bellitti G., Fedeli M. (a cura di) (2024), *Digitalizzazione, interoperabilità e intelligenza artificiale*, Roma, Istat

Bellitti G., Fedeli M. (a cura di) (2023), *Trasformazione digitale della Pubblica amministrazione. Metodi per l'interoperabilità per lo sviluppo di e-service*, Roma, Istat

Berners-Lee T., Hendler J., Lassila O. (2001), The Semantic Web, *Scientific American*, 284, n.5, pp.34-43 <DOI:10.1038/scientificamerican0501-34>

Bizer C., Heath T., Berners-Lee T. (2023), Linked data: The story so far, in Seneviratne O., Hendler J. (eds.), *Linking the World's Information: Essays on Tim Berners-Lee's Invention of the World Wide Web*, New York, ACM, pp.115-143

Calvanese D., De Giacomo G., Lembo D., Lenzerini M., Poggi A., Rodriguez-Muro M., Rosati R., Ruzzi M., Savo D.F. (2011), The MASTRO system for ontology-based data access, *Semantic Web*, 2, n.1, pp.43-53

Ceccarelli C. (2025), La metodologia per la gestione e produzione di informazione statistica, in Bellitti G., Ceccarelli C. (a cura di), *Pratiche per il trattamento delle informazioni nel processo di produzione statistica: compliance e metodi*, Roma, Istat, pp.27-35

Commissione europea (2024), *Interoperable Europe Act*, Regulation n. 903/2024, Bruxelles, European Union

Commissione europea (2022), *Final evaluation of the European Interoperability Framework (EIF)*, Staff Working Document n.720/2022, Bruxelles, European Union

Commissione europea (2020), *A European Strategy for Data*, Communication n.66/2020, Bruxelles, European Union

Commissione europea (2017), *European Interoperability Framework – Implementation Strategy*, Communication n.134/2017, Bruxelles, European Union

Commissione europea (2010), *Towards interoperability for European public services***,** Communication n.744/2010, Bruxelles, European Union

Davide F. (2024), Semic 2024: verso un'Europa digitale e interoperabile, *agendadigitale.eu*, 21 giugno

Dawes S.S. (1996), Interagency information sharing: Expected benefits, manageable risks, *Journal of Policy Analysis and Management*, 15, n.3, pp.377-394

De Seta C. (1985), Principi giurisprudenziali in tema di buon andamento, *Amministrazione e Contabilità dello Stato e degli enti pubblici*, n.4, p.27 ss.

Dietrich D. (2012), Linked Data, *ePSIplatform Topic Report* n.11

Eurostat (2019), *ESS Handbook for Quality Reports*, European Union

Euzenat J., Shvaiko P. (2013), *Ontology Matching*, Berlin Heidelberg, Springer-Verlag

Galetta D.U. (2021), Digitalizzazione e diritto ad una buona amministrazione, *Rivista Interdisciplinare sul Diritto delle Amministrazioni Pubbliche*, n.3, pp.197-205

Gruber T. (1995), Toward principles for the design of ontologies used for knowledge sharing, *International Journal of Human-Computer Studies*, 43, n.5-6, pp.907-928

Guarino N. (1998), Formal Ontology and Information Systems, in Guarino N. (ed.), *Proceedings of FOIS*'98, Amsterdam, Verlag, IOS Press, pp.3-15

Harpring P. (2010), *Introduction to controlled vocabularies: terminology for art, architecture, and other cultural works*, Los Angeles, Getty Research Institute

Hendler J., Gandon F., Allemang D. (2020), *Semantic web for the working ontologist: Effective modeling for linked data, RDFS, and OWL*, San Rafael, Morgan & Claypool

Inapp, Chiurco F., Cioccolo V., Sofronic B. (2025), *Verso l'interoperabilità. Una sperimentazione di interconnessione tra strumenti e piattaforme ICT dell'Inapp*, Inapp WP n.138, Roma, Inapp

ISO (2011), *ISO 25964-1: Information and documentation. Thesauri and interoperability with other vocabularies*, Geneva, ISO

ISO/IEC (2022), *ISO/IEC 27002:2022: Information security, cybersecurity and privacy protection – Information security controls*, Geneva, ISO and IEC

Istat (2023), *Censimento permanente delle Istituzioni Pubbliche*, Roma, Istat

Istat (2022), *Rapporto annuale 2022. La situazione del Paese*, Roma, Istat

Istat (2021), *Il sistema di documentazione dei dati amministrativi in Istat*, Roma, Istat

Lenzerini M. (2023), Data Modeling and AI: From Semantic Networks to Knowledge Graphs, *12th International Conference on Data Science, Technology and Applications,* Rome, July 2023

Lenzerini M. (2011), Ontology-based data management, *Proceedings of the 20th ACM international conference on Information and knowledge management (CIKM '11)*, Association for Computing Machinery, New York

Meijer A. (2014), Transparency, in Bovens M., Goodin R.E., Schillemans T. (eds.), *The Oxford Handbook of Public Accountability*, Oxford, Oxford University Press, pp.507-524

OECD (2019), *The Path to Becoming a Data-Driven Public Sector*, OECD Digital Government Studies, Paris

Polizzi M., Letardi S. (2023), Il Catalogo Nazionale Dati: far parlare i dati per innovare i servizi, *Conferenza Italiana Scienze Regionali*, Napoli, Settembre 2023

Puglisi P., Radini R. (2025), Casi di contribuzione al Catalogo, in Bellitti G., Fedeli M. (a cura di), *Catalogo Nazionale Dati. Metodologia per l'interoperabilità semantica*, Roma, Istat, pp.23-26

Ranatunga S., Ødegård R.S., Jetlund K., Onstein E. (2025), Use of Semantic Web Technologies to Enhance the Integration and Interoperability of Environmental Geospatial Data: A Framework Based on Ontology-Based Data Access, *ISPRS International Journal of Geo-Information*, 14, n.2, n.52

Regolamento (UE) 2016/679 del Parlamento europeo e del Consiglio del 27 aprile 2016 relativo alla protezione delle persone fisiche con riguardo al trattamento dei dati personali, nonché alla libera circolazione di tali dati e che abroga la direttiva 95/46/CE (regolamento generale sulla protezione dei dati) , in Gazzetta Ufficiale dell'Unione europea, L 119 del 4 maggio 2016, pp.1-88

Rivas L., Crowley J. (2018), *Using Administrative Data to Enhance Policymaking in Developing Countries: Tax Data and the National Accounts*, IMF Working Papers 2018/175, Washington, International Monetary Fund

Spasiano M.R. (2017), Il principio di buon andamento, in Sandulli M.A. (a cura di), *Principi e regole dell'azione amministrativa*, Milano, Giuffrè, pp.121-144

Stiglitz J.E. (1999), Knowledge as a global public good, in Kaul I., Grunberg I., Stern M. (eds.), *Global public goods: International cooperation in the 21st century,* Oxford, Oxford University Press, pp.308-325

Studer R., Benjamins V.R., Fensel D. (1998), Knowledge engineering: Principles and methods, *Data & knowledge engineering*, 25, n.1-2, pp.161-197

Ubaldi B. (2013), *Open Government Data: Towards Empirical Analysis of Open Government Data Initiatives*, OECD Working Papers on Public Governance No. 22, Paris, OECD Publishing

UNECE (2021), *Guidelines for Assessing the Quality of Administrative Sources for Use in Censuses*, Geneva, United Nations

UNECE (2011), *Using Administrative and Secondary Sources for Official Statistics. A Handbook of Principles and Practices*, New York and Geneva, United Nations

W3C (2024), *Data Catalog Vocabulary (DCAT) – Version 3*, W3C Recommendation, 22

W3C (2009), *SKOS Simple Knowledge Organization System Reference*, W3C Recommendation, 18 August 2009

Wallgren A., Wallgren B. (2007), *Register-based statistics: Administrative data for statistical purposes*, Chichester, Wiley & Sons

Wilkinson M., Dumontier M., Aalbersberg I. *et al.* (2016), The FAIR Guiding Principles for scientific data management and stewardship, *Sci Data*, 3, 160018

Zaytseva K., Ďurčo M. (2020), *Controlled Vocabularies and SKOS. Version 1.1.0,* DARIAH Campus

Zuiderwijk A., Chen Y., Salem F. (2021), Implications of the use of artificial intelligence in public governance: A systematic literature review and a research agenda, *Government Information Quarterly*, 38, n.3

Alessio Buonocore
alessio.buonocore@istat.it
Collaboratore di ricerca presso la Direzione centrale per la metodologia e il disegno dei processi statistici (DCME) dell'Istituto nazionale di statistica (ISTAT). Partecipa alle attività del Catalogo nazionale dati, curando la modellazione e lo sviluppo di risorse semantiche, contribuendo alla definizione di modelli concettuali e alla standardizzazione delle strutture informative.

Eugenio Concetti
eugenio.concetti@istat.it
Collaboratore di ricerca presso la Direzione centrale per la metodologia e il disegno dei processi statistici (DCME) dell'Istituto nazionale di statistica (ISTAT). Con competenze in indagini statistiche e censuarie, si occupa di revisioni semantiche nel National Data Catalog. Membro dell'Associazione italiana di terminologia, ha collaborato con istituzioni culturali, festival e associazioni come responsabile organizzativo e di comunicazione.

Giampiero D'Alessandro
g.dalessandro@istat.it
Ricercatore presso la Direzione centrale per la metodologia e il disegno dei processi statistici (DCME) dell'Istituto nazionale di statistica (ISTAT). Per il Catalogo nazionale dati si occupa di modellazione di ontologie e realizzazione di vocabolari controllati. Tra le pubblicazioni, in corso di stampa, più recenti: *Terminology and Conceptualization* e *Ontology and Terminology Complementarity*.

Adriana Fidanza
adriana.fidanza@istat.it
Collaboratrice di ricerca presso la Direzione centrale per la metodologia e il disegno dei processi statistici (DCME) dell'Istituto Nazionale di Statistica (ISTAT). Per il Catalogo Nazionale Dati si occupa della modellazione di ontologie, dello sviluppo di vocabolari controllati, del monitoraggio delle risorse semantiche e della sperimentazione di soluzioni basate sull'intelligenza artificiale.

Stefania Pantoni
stefania.pantoni@istat.it
Collaboratrice di ricerca presso la Direzione centrale per la metodologia e il disegno dei processi statistici (DCME) dell'Istituto nazionale di statistica (ISTAT). Nell'ambito del progetto NDC (National Data Catalog), finanziato grazie ai fondi PNRR (Missione 1-Componente-1 Investimento 1.3: Dati e Interoperabilità), supporta le attività di analisi e definizione concettuale del dominio di interesse, necessarie alla creazione degli asset semantici – ontologie, vocabolari controllati e schemi di dati – che abilitano l'interoperabilità tra sistemi informativi della Pubblica amministrazione.

Roberta Radini
radini@istat.it
Primo Ricercatore presso l'Istat, lavora nella direzione delle Metodologie nel servizio architettura integrata dei dati e dei processi. È responsabile del team che si occupa di modellazione ontologica e architettura informativa del Sistema integrato dei registri (SIR) dell'Istat. Attualmente coordina il Team Istat di Sviluppo metodologie per l'interoperabilità semantica nell'ambito del progetto National Data Catalog (NDC). In questo contesto è responsabile della formazione, del supporto e all'affiancamento delle Pubbliche amministrazioni nell'adozione e nell'utilizzo del NDC.

Open data e orientamento delle politiche pubbliche nella politica di coesione

Dall'accountability al policy learning

Giorgia Marinuzzi
IFEL- Fondazione ANCI

Walter Tortorella
IFEL- Fondazione ANCI

L'articolo analizza il ruolo degli open data nella politica di coesione europea, sostenendo il passaggio dalla funzione di mera trasparenza e accountability a quella di infrastruttura per il policy learning. Attraverso il caso di OpenCoesione e l'analisi dei dati FESR e FSE+ 2021-2027, evidenzia come interoperabilità, tempestività e standardizzazione dei dati possano supportare allocazione delle risorse, adattamento delle politiche e apprendimento istituzionale. Gli open data emergono così come strumenti strategici per una governance più efficace, flessibile e orientata ai risultati.

Open data and the orientation of public policies in cohesion policy. From accountability to policy learning

his article examines the role of open data in European Cohesion Policy, arguing for a shift from transparency and accountability functions towards a policy-learning infrastructure. Drawing on the Italian OpenCoesione experience and the analysis of ERDF and ESF+ 2021-2027 data, it shows how interoperability, timeliness and data standardization can support resource allocation, policy adaptation and institutional learning. Open data therefore emerge as strategic tools for a more effective, flexible and results-oriented governance of cohesion policy.

DOI: 10.53223/Sinappsi_2026-02-04

Citazione
Marinuzzi G., Tortorella W. (2026), Open data e orientamento delle politiche pubbliche nella politica di coesione. Dall'accountability al policy learning, *Sinappsi*, XVI, n.2, pp.65-76

Parole chiave
Interoperabilità
Open data
Politiche pubbliche

Keywords
Interoperability
Open data
Public policies

1. Open data, politica di coesione e apprendimento delle politiche pubbliche

Negli ultimi quindici anni la disponibilità di dati aperti ha progressivamente modificato il rapporto tra trasparenza amministrativa, accountability e analisi delle politiche pubbliche. Nell'ambito della politica di coesione europea, tale evoluzione ha trovato in Italia un'espressione particolarmente significativa attraverso OpenCoesione, iniziativa di open government che dal 2012 rende disponibili dati di dettaglio sui progetti finanziati con risorse europee e nazionali.

La letteratura ha evidenziato come OpenCoesione abbia rappresentato non soltanto uno strumento di trasparenza verso cittadini e stakeholder, ma anche una piattaforma di sperimentazione per il monitoraggio civico, il riuso degli open data e l'analisi delle politiche pubbliche. Diversi contributi hanno infatti utilizzato i dati di OpenCoesione per studiare forme di partecipazione civica e accountability (Reggi e Dawes 2016), esperienze di monitoraggio civico quali 'Monithon'[1] e 'A Scuola di OpenCoesione' (Collesi e De Luca 2022; Buttiglione e Reggi 2015), nonché aspetti territoriali e valutativi degli investimenti della politica di coesione (Albanese e De Blasio 2018; Fusacchia *et al.* 2024).

1 Cfr. https://opencoesione.gov.it/it/monithon/ .

In questo quadro, il presente contributo non si propone di valutare l'efficacia degli interventi finanziati, bensì di riflettere sul ruolo degli open data come infrastruttura informativa a supporto dell'orientamento delle politiche pubbliche. L'obiettivo è mostrare come la disponibilità di dati aperti e interoperabili possa favorire non soltanto la rendicontazione degli investimenti, ma anche processi di apprendimento istituzionale, riallocazione delle risorse e adattamento delle politiche nel corso della loro attuazione.

Per rendere più concreto questo argomento, l'analisi si concentra sui Fondi FESR e FSE+ del ciclo di programmazione 2021-2027. La scelta è motivata dal fatto che tali Fondi rappresentano i principali strumenti finanziari della politica di coesione europea e coprono tipologie di intervento profondamente differenti: investimenti infrastrutturali, innovazione e competitività nel caso del FESR; occupazione, istruzione, inclusione sociale e capitale umano nel caso del FSE+. L'eterogeneità delle loro finalità consente di evidenziare le potenzialità degli open data nell'analisi di politiche pubbliche caratterizzate da beneficiari, modalità attuative e risultati attesi molto diversi tra loro.

2. Dagli elenchi dei beneficiari agli open data

C'erano una volta gli 'elenchi dei beneficiari' dei Fondi strutturali. File, per lo più in pdf, contenenti liste lunghissime di progetti ascrivibili a Programmi operativi regionali o nazionali con l'informazione relativa all'Asse di riferimento, alla denominazione dei soggetti beneficiari e all'importo finanziario.

Era il 2010 quando ci venne l'idea di mettere in produzione la prima delle 15 edizioni del volume *La dimensione territoriale della politica di coesione* di IFEL[2]-Fondazione ANCI, con l'obiettivo di analizzare l'attuazione del Fondi strutturali in Italia fino a comprendere il grado di coinvolgimento dei comuni in qualità di beneficiari. Per fare ciò era necessario mappare tutti i progetti esistenti e per ciascuno di essi risalire al beneficiario.

Nel 2010, in pieno ciclo di programmazione 2007-2013, le Autorità di Gestione erano tenute a pubblicare tali informazioni all'interno dei cosiddetti elenchi dei beneficiari, ma il confine tra semplice adempimento dell'Amministrazione titolare ed effettiva possibilità di permettere a cittadini, ricercatori, politici di utilizzare quel dato, più che una linea assomigliava ad un abisso:

- non esisteva un unico luogo in cui gli elenchi dei beneficiari di tutti i Programmi venissero raccolti. Per consultare i dati era necessario cercare l'elenco specifico nei diversi siti web dei singoli Programmi Operativi;
- non erano previste date definite e condivise di rilascio degli elenchi, inficiando la possibilità di confrontarli poiché riferiti a tempi diversi;
- non erano previste strutture standard dei file, né in termini di contenuti (variabili diverse tra file), né in termini di organizzazione delle informazioni (ordine diverso di esposizione dei dati nelle tabelle);
- non era prevista alcuna categorizzazione dei soggetti beneficiari associati ai singoli progetti;
- non era previsto un sistema di controllo automatico delle informazioni contenute nei file. Ricordiamo ancora il file pubblicato da un POR delle regioni cosiddette 'Competitività', coincidenti al tempo con il Centro-Nord geografico, che, per un errore umano, riportava come prima riga di ciascun foglio lo stesso progetto replicato tante volte quante erano le pagine dell'elenco;
- il formato dei file rilasciati avveniva nella gran parte dei casi in pdf (pochissime eccezioni di formati elaborabili come Excel), in un periodo storico in cui i software di lettura OCR dei pdf si trovavano nella biblioteca dell'Istat e non erano diffusi tra i ricercatori. Spesso si trattava addirittura di scansioni di fotocopie, la soluzione peggiore in termini di uso e riuso dei dati.

Del resto, non si trattava di un'inefficienza o di un'anomalia italiana, ma era proprio il Regolamento generale sui Fondi Strutturali 2007-2013 (n.1083/2006 art. 69) che richiedeva alle Autorità di Gestione dei Programmi di pubblicare sul proprio sito web 'soltanto' un insieme minimo di informazioni sui beneficiari dei Fondi, sulle operazioni finanziate e sul relativo importo.

In 15 anni di redazione di un volume di ricerca dedicato alla dimensione territoriale delle politiche di coesione possiamo affermare che di passi in avanti ne abbiamo visti, e con il Regolamento generale sui Fondi strutturali 2014-2020 (n. 1303/2013 art. 115) si è acceso un faro sul tema dell'accessibilità ai dati relativi all'attuazione della politica di coesione. Il Regolamento richiedeva agli Stati membri, anche attraverso un portale web unico, di pubblicare le informazioni sui beneficiari in formato aperto (con licenza di riuso e metadati) ed elencava il livello minimo di informazioni da pubblicare con maggiori dettagli rispetto al periodo 2007-2013.

Nel 2012 nasce OpenCoesione[3], l'iniziativa di open government sulle politiche di coesione in Italia

2 Istituto per la Finanza e l'economia locale, cfr. https://www.fondazioneifel.it/.

3 Cfr. https://opencoesione.gov.it/it/.

coordinata dall'attuale Dipartimento per le Politiche di coesione della Presidenza del Consiglio dei Ministri, individuato come il portale unico italiano per l'attuazione della politica di coesione, in grado di offrire una banca dati sistematizzata di tutti i progetti dei singoli Fondi europei con l'indicazione delle variabili finanziarie, dei soggetti che a vario titolo (programmatore, attuatore, beneficiario) sono coinvolti, della localizzazione e dello stato di avanzamento del progetto. L'obiettivo di OpenCoesione è infatti favorire un migliore uso delle risorse pubbliche attraverso la diffusione sul portale e il riutilizzo di dati e informazioni sugli interventi finanziati con risorse nazionali ed europee.

Il Regolamento generale sui Fondi strutturali 2021-2027 (n. 1060/2021 art. 46) conferma agli Stati membri la richiesta di prevedere un portale web unico che offra accesso a dati sui programmi cofinanziati, dettagliando all'art. 49 le informazioni, i formati e la frequenza minima di aggiornamento richiesti. Anche per il periodo di programmazione in corso OpenCoesione rappresenta il portale unico italiano sull'attuazione delle politiche di coesione.

Grazie alle indicazioni contenute nel Regolamento n. 1060 del 2021, il cambio di rotta è epocale perché si superano tutti i limiti elencati prima: esiste un unico luogo dal quale scaricare l'elenco dei progetti e dei relativi beneficiari per tutti i Programmi e per tutti i Fondi; i dati sono rilasciati in formato csv, che si caratterizza per la sua semplicità d'uso, compatibilità universale e leggerezza; infine sono previsti aggiornamenti periodici dell'intera base dati che al proprio interno contiene informazioni riferite ad una medesima data.

Finalmente si può parlare di open data, nel vero senso della parola di 'dati aperti', ossia informazioni pubblicate in formato tale che siano accessibili, soprattutto via internet, senza limitazioni secondo l'identità o lo scopo dell'utente; dati che siano elaborabili da un'applicazione informatica senza la necessità di un software specifico e che siano accompagnati da licenze che garantiscano libertà nell'uso e nel riuso.

La letteratura sugli open data ha chiarito, tuttavia, che l'apertura non produce automaticamente impatto. Janssen *et al.* (2012) mostrano come benefici attesi e realtà siano mediati da barriere (qualità, standard, capacità di riuso, incentivi). In breve: i dataset da soli non cambiano le politiche pubbliche ma possono orientarle. Un passaggio cruciale è la *targeted transparency*: la trasparenza funziona quando è disegnata per rendere comparabili e azionabili informazioni che modificano comportamenti decisionali (di amministrazioni, stakeholder, cittadini). Applicata alla coesione, questa impostazione implica che l'open data utile non è quello che descrive tutto, ma quello che consente di rispondere a domande operative: Stiamo spendendo e realizzando nei tempi coerenti con gli obiettivi? Dove si accumulano colli di bottiglia e perché? Quali territori restano sistematicamente indietro rispetto ai fabbisogni? Quali strumenti funzionano meglio in relazione a quali contesti?

L'attuale ciclo di programmazione 2021-2027 è strutturato attorno all'Accordo di Partenariato come documento di orientamento strategico nazionale e ai Programmi (nazionali e regionali) che declinano obiettivi e risorse. Analogamente lo erano in qualche modo anche i precedenti cicli di programmazione, che però facevano della raccolta dati più un elemento riempitivo adempimentale che una cartina di tornasole finalizzata a massimizzare il contributo della coesione riorientando di volta in volta le scelte alle priorità correnti ed emergenti e ad aumentare l'impatto, introducendo maggiore flessibilità e riallineamento degli investimenti[4]. Finalmente, con il monitoraggio di OpenCoesione, che negli anni è diventato sempre più fruibile e accurato, gli open data finiscono per servire a tre funzioni proprie delle politiche pubbliche, diverse dalla mera accountability: allocazione informata (provare a orientare dove investire di più/meno); adattamento (correzioni in itinere su strumenti e implementazione); apprendimento comparativo (capire cosa funziona, per chi, e in quali condizioni). Costruire una catena del valore che dal dato arrivi alla decisione è la principale prova di maturazione del contesto italiano, e anche il rinnovamento della sezione open data di OpenCoesione come catalogo più descrittivo e filtrabile segnala che la usabilità del dato viene trattata come requisito e non come dettaglio.

3. L'interoperabilità dei dati amministrativi

La possibilità di monitorare e valutare la politica di coesione attraverso questa nuova stagione di apertura dei dati è sicuramente un'opportunità concreta e positiva per i ricercatori; tuttavia, è sempre necessario leggere e analizzare gli open data con spirito critico e prudenza, in quanto non sono esenti da errori, omissioni o possibili distorsioni che, se non intercettati, portano ad una moltiplicazione esponenziale e incontrollabile dell'anomalia.

Nel caso di OpenCoesione, uno dei temi di attenzione riguarda ad esempio la classificazione dei soggetti beneficiari che, se non ben controllata,

4 Cfr. Communication from the Commission to the European Parliament and the Council (2025), *A modernised Cohesion policy: The mid-term review*, COM(2025) 163 final.

può generare attribuzioni errate. Il sistema contiene, infatti, una 'pre-categorizzazione' dei soggetti che però può essere affinata e modellizzata in base alle esigenze di analisi direttamente connesse con la domanda di ricerca alla quale si vuole rispondere. Nel caso di IFEL, ad esempio, per l'obiettivo che ci siamo posti sin dalla prima delle 15 edizioni del volume *La dimensione territoriale della politica di coesione*, ossia comprendere il grado di coinvolgimento dei comuni nell'utilizzo delle risorse dei Fondi strutturali europei, era cruciale individuare tutti i comuni beneficiari di tali risorse anche nel caso in cui nel campo della denominazione del beneficiario non venisse esplicitata la parola 'comune' o 'amministrazione comunale'. Per fare ciò è possibile 'incrociare' i codici fiscali dei soggetti beneficiari contenuti nella banca dati OpenCoesione con i codici fiscali degli enti presenti nell'Anagrafe degli Enti della PA, una banca dati di BDAP[5]-MEF che raccoglie e integra i dati anagrafici provenienti da differenti fonti istituzionali, come:

- SIOPE - Sistema informativo sulle operazioni degli enti pubblici;
- IPA - Indice delle Pubbliche amministrazioni;
- MIM - Ministero dell'Istruzione e del merito;
- ISTAT - Registro statistico delle Amministrazioni pubbliche (c.d. 'Lista S13';
- DT - Dipartimento del Tesoro del Ministero dell'Economia e delle finanze.

Si precisa che per i beneficiari individuali i codici fiscali sono anonimizzati.

Volendo restare nel tema dell'interoperabilità dei dati amministrativi, si segnalano altri esempi di possibili collegamenti utili ai fini della ricerca: un esempio è quello di 'agganciare' ai beneficiari comunali, sempre tramite BDAP, il codice identificativo univoco del comune che viene assegnato dall'Istat al singolo ente, un codice a sei cifre che combina il numero progressivo della provincia con quello del comune. In questo modo si crea un ponte con tutte le principali banche dati istituzionali, Istat e non, che adottano tale codifica per rappresentare i dati comunali. Ed è così, ad esempio, che nel caso in cui si vogliano valutare le risorse provenienti da Fondi strutturali impiegate per un determinato servizio si riesce anche a risalire alla relativa domanda potenziale che potrebbe riguardare una fascia specifica di popolazione.

Il codice Istat comunale è invece già esplicitato nei db OpenCoesione nella sezione relativa alla localizzazione degli interventi, quindi, a prescindere dal coinvolgimento dei comuni nel ruolo di beneficiari: in questo modo è possibile dunque studiare l'impatto della politica di coesione sui territori, valutando le differenze di variabili macro nel periodo pre e post erogazione delle risorse o tra aree che hanno ricevuto o meno Fondi. Si segnala che soprattutto nel caso di interventi riguardanti la mobilità/viabilità (es. lavori stradali) può essere indicata una 'plurilocalizzazione' che non permette di risalire alla quota specifica di risorse ricadenti in un determinato territorio.

Un'altra possibilità relativa all'interoperabilità dei db, sempre partendo da quelli di attuazione della politica di coesione, ha a che fare con la mappatura delle diverse fonti di finanziamento di un progetto, come ad esempio le risorse PNRR, che vengono monitorate sul portale Italia Domani[6] che 'pesca' direttamente dal sistema di monitoraggio e rendicontazione ReGiS del MEF. Per ciascun progetto, infatti, è possibile generare un codice univoco concatenando il Codice locale progetto (CLP) (già di per sé univoco in OpenCoesione) con il Codice unico di progetto (CUP) (che invece presenta repliche): non basta attenersi al CLP perché nei db Italia Domani i codici locali progetto presentano invece duplicazioni.

La tabella 1, di seguito, riporta un quadro sinottico con le principali caratteristiche in termini di accessibilità, copertura temporale e spaziale, dettaglio analitico, punti di forza e debolezza delle banche dati fin qui nominate, ossia OpenCoesione, ReGiS e Italia Domani.

Se, da un lato, è auspicabile che gli open data consentano progressivamente di superare una lettura esclusivamente finanziaria — centrata su quanto si spende — per adottare un approccio *problem-driven*, orientato a individuare dove si concentrano i bisogni e in che modo le risorse riescono a intercettarli, dall'altro è fondamentale che i diversi sistemi di monitoraggio siano tra loro interoperabili. Una volta definite priorità e finalità specifiche, la possibilità di ricondurre progetti e risorse a tali obiettivi diventa infatti uno strumento di riprogrammazione in itinere. Il confronto con indicatori territoriali, anch'essi costruiti a partire da microdati, può così alimentare una discussione pubblica più consapevole sulle scelte allocative e sulla loro eventuale revisione. L'obiettivo è evitare che le risorse si concentrino nei territori in cui la capacità di assorbimento è maggiore, anziché in quelli in cui i bisogni risultano più intensi.

Con dati aggiornati e storicizzati, si possono costruire segnali anticipatori: stalli, ritardi ricorrenti, pattern di pagamento anomali, concentrazione di

5 Cfr. Banca dati amministrazioni pubbliche, https://openbdap.rgs.mef.gov.it/it/Home/AnagrafeEnti.

6 Cfr. https://www.italiadomani.gov.it/content/sogei-ng/it/it/home.html.

Tabella 1. Le principali caratteristiche delle banche dati OpenCoesione, ReGiS e Italia Domani

CARATTERISTICHE	BANCA DATI		
	OpenCoesione	ReGiS (Sistema nazionale monitoraggio PNRR e coesione 2021-2027)	Italia Domani Open Data
Accessibilità	Open data scaricabili (csv), API, dashboard pubbliche	Accesso amministrativo; apertura parziale tramite dataset pubblicati su Italia Domani e iniziative civiche	Open data scaricabili (csv, json, xlsx)
Copertura temporale	2000-2006, 2007-2013, 2014-2020, 2021-2027	Dal 2021 in avanti	Dal 2021
Copertura spaziale	Nazionale, regionale, comunale, singolo progetto	Nazionale fino al singolo progetto	Nazionale, regionale, comunale, progetto
Dettaglio analitico	Molto elevato: CUP, CLP, beneficiari, soggetti attuatori, localizzazione, avanzamento finanziario e procedurale	Molto elevato: milestone, target, procedure, gare, pagamenti, cronoprogrammi	Medio-alto: CUP, soggetti, localizzazione, misure PNRR
Punti di forza	Massima trasparenza, serie storica lunga, dati progetto-level, interoperabilità con CUP e dati territoriali	Sistema gestionale ufficiale; integra PNRR e progressivamente monitoraggio coesione 2021-2027	Elevata accessibilità; aggiornamenti frequenti
Principali limiti	Limitata disponibilità di output e outcome; qualità dipendente dal monitoraggio amministrativo; dati 2021-2027 inizialmente raccolti extra-sistema	Non completamente open; accesso differenziato; struttura complessa; storicità limitata	Meno ricco di ReGiS; focalizzato sul PNRR; minore profondità storica

Fonte: IFEL-Area Studi e statistiche territoriali, 2026

criticità per misura/soggetto attuatore/beneficiari/Fondo. Questo è esattamente ciò che serve quando la *mid-term review* spinge verso riallocazioni e correzioni ricorrenti – come avviene in parte nelle politiche di coesione – per aumentarne l'impatto. Gli open data rendono possibile, allora, il confronto sistematico tra strumenti simili in contesti diversi: tempi medi, varianti, performance implementativa, risultati proxy. Questa capacità comparativa è cruciale anche nel dibattito post-2027, dove una parte della discussione riguarda proprio come rendere la politica di coesione più efficace e più agile rispetto ad altri strumenti (in primis il confronto è con modelli più *performance-oriented* quale il PNRR). Tre elementi emergono come condizioni abilitanti:

1. Tempestività e continuità di pubblicazione: senza dati aggiornati (e storico/versionamento), l'orientamento si trasforma in archeologia amministrativa. Sul ciclo 2021-2027, ad esempio, alcune comunicazioni pubbliche indicano che la pubblicazione dei dati progetto può dipendere dalla piena operatività del nuovo sistema nazionale di monitoraggio, con effetti sulla disponibilità tempestiva.
2. Standard e metadati: sono necessari per rendere comparabili programmi e progetti (e quindi valutabili), evitando che il dato aperto diventi *'open but unusable'*.
3. Integrazione con cicli decisionali: il dato deve attivare routine (verifiche di portafoglio, revisioni periodiche, log delle decisioni), altrimenti resta consultazione fine a se stessa se non oggetto di strumentalizzazione.

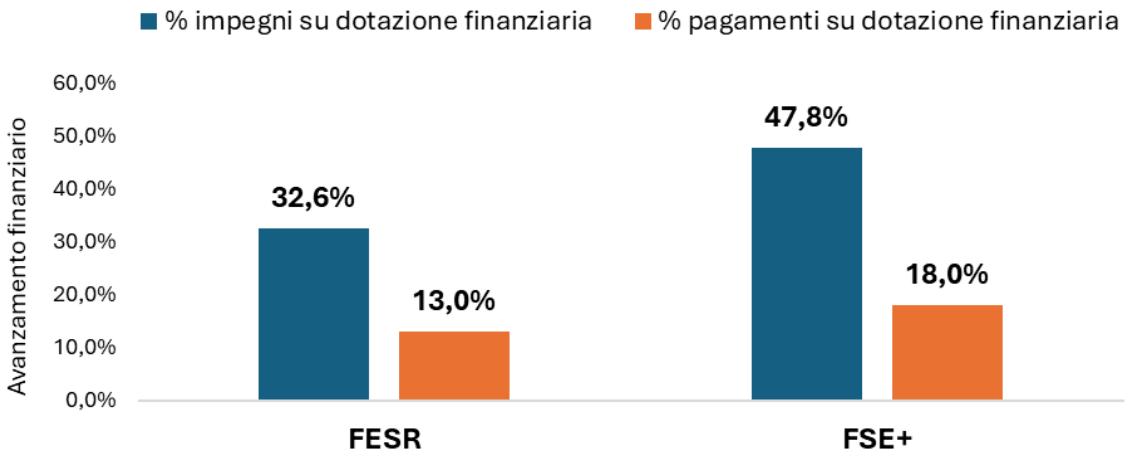
4. L'attuazione dei progetti FESR e FSE+ 2021-2027

I Fondi principali della politica di coesione 2021-2027, il Fondo europeo di sviluppo regionale (FESR) e il Fondo sociale europeo Plus (FSE+), in Italia hanno una dotazione finanziaria pari a 44,0 miliardi di euro e 28,6 miliardi di euro rispettivamente.

Il dato di avanzamento finanziario viene rilasciato ad intervalli bimestrali dal MEF-IGRUE attraverso dei bollettini di monitoraggio[7] che riportano informazioni aggregate per Fondo, per categorie di regioni (più sviluppate, in transizione, meno

7 Cfr. https://www.rgs.mef.gov.it/VERSIONE-I/attivita_istituzionali/monitoraggio/rapporti_finanziari_ue/monitoraggio_politiche_di_coesione_2021-2027_e_2014-2020/.

Figura 1. Lo stato di avanzamento finanziario di FESR e FSE+ 2021-2027 alla data del 28 febbraio 2026



Fonte: elaborazione IFEL-Area Studi e statistiche territoriali su dati MEF, 2026

sviluppate), per Programma nazionale e regionale sui livelli di impegni e di spesa raggiunti. Non si tratta di open data.

L'ultimo bollettino disponibile al momento della redazione del presente articolo fa riferimento al 28 febbraio 2026, e riporta un dato di avanzamento finanziario leggermente più elevato nel caso del FSE+ rispetto al FESR (figura 1): nel primo caso, infatti, gli impegni si attestano al 47,8% della dotazione del Fondo e i pagamenti al 18,0%; nel caso del FESR, invece, gli impegni hanno raggiunto il 32,6% e i pagamenti il 13,0% del valore complessivo del Fondo.

I livelli di attuazione italiani sono confrontabili con quelli degli altri Stati membri attraverso la piattaforma Cohesion Open Data[8] della Commissione europea. In questo caso si tratta di dati aperti, scaricabili in formato csv o xls, ma con un livello di dettaglio piuttosto 'alto', ossia informazioni disponibili per Paese, Fondo, categoria di regioni, Programma e Obiettivo Specifico. Il dato non è disponibile a livello di progetto.

Come prima anticipato, lo scenario cambia radicalmente con i DB di OpenCoesione, in cui l'unità di rilevazione è il singolo progetto per il quale si dispone di un set di informazioni relative alla natura e alla descrizione dell'intervento stesso, alla sua localizzazione, ai soggetti coinvolti nelle diverse fasi di realizzazione del progetto, alle risorse finanziarie in campo e all'avanzamento procedurale.

La base dati OpenCoesione è alimentata dalle informazioni provenienti dal Sistema nazionale di monitoraggio gestito dall'Ispettorato generale per i Rapporti finanziari con l'Unione europea (IGRUE) del Ministero dell'Economia e delle finanze per i cicli di programmazione 2000-2006, 2007-2013 e 2014-2020. Tuttavia, per quanto concerne i progetti finanziati con risorse europee e nazionali del ciclo 2021-2027, la fonte deriva da una rilevazione apposita condotta da MEF-IGRUE, di cui vengono pubblicati i dati riferiti a progetti dotati del Codice unico di progetto.

Dunque, quali informazioni si possono trarre dall'elaborazione e dall'analisi degli open data italiani su FESR e FSE+? Secondo il quadro aggiornato al 31 dicembre 2025, per il ciclo di programmazione 2021-2027 si contano oltre 39 mila progetti FESR (tabella 2) e 76 mila FSE+ (tabella 3).

È possibile risalire alle tematiche principali sulle quali si concentrano i progetti: nel caso del FESR prevalgono la competitività delle imprese (oltre la metà degli interventi) e la ricerca/innovazione (21,8%). I progetti FSE+, in termini numerici, si dividono principalmente tra istruzione, formazione, occupazione e lavoro.

L'analisi per tema dei progetti cambia facendo riferimento ai costi rendicontabili UE[9]. Dei 22,27 miliardi di euro associati ai progetti FESR la quota destinata alla competitività delle imprese sfiora il 32% e aumenta il peso di tematiche come quelle ambientali e quelle connesse alla capacità amministrativa, che superano il 10% delle risorse. I

8 Cfr. https://cohesiondata.ec.europa.eu/.

9 La variabile "Costo rendicontabile UE" si riferisce al costo ammesso al contributo comunitario, ossia le risorse che possono essere considerate per la richiesta dei rimborsi alla Commissione europea. Il "Costo rendicontabile UE" comprende la quota comunitaria e il corrispondente cofinanziamento nazionale obbligatorio.

Tabella 2. I progetti FESR 2021-2027, per tema d'intervento, 31 dicembre 2025

Tema	N. progetti		Costo rendicontabile UE	
	v.a.	% sul totale	v.a. (euro)	% sul totale
Ambiente	1.561	4,0	2.817.480.867	12,6
Capacità amministrativa	1.271	3,3	2.245.345.812	10,1
Competitività delle imprese	20.438	52,3	7.085.267.643	31,8
Cultura e turismo	662	1,7	725.826.739	3,3
Energia	1.215	3,1	335.438.189	1,5
Inclusione sociale e salute	1.504	3,8	1.817.482.453	8,2
Istruzione e formazione	2.000	5,1	1.222.178.589	5,5
Occupazione e lavoro	170	0,4	39.895.117	0,2
Reti e servizi digitali	1.076	2,8	1.504.592.848	6,8
Ricerca e innovazione	8.533	21,8	1.985.515.969	8,9
Trasporti e mobilità	671	1,7	2.494.476.262	11,2
Totale	**39.101**	**100,0**	**22.273.500.488**	**100,0**

Fonte: elaborazione IFEL-Area Studi e statistiche territoriali su dati OpenCoesione aggiornati al 31.12.2025

Tabella 3. I progetti FSE+ 2021-2027, per tema d'intervento, 31 dicembre 2025

Tema	N. progetti		Costo rendicontabile UE	
	v.a.	% sul totale	v.a. (euro)	% sul totale
Capacità amministrativa	565	0,7	1.321.599.975	9,3
Competitività delle imprese	69	0,1	209.746.552	1,5
Cultura e turismo	13	0,02	86.606.597	0,6
Inclusione sociale e salute	6.314	8,2	3.772.283.970	26,6
Istruzione e formazione	37.910	49,5	4.055.787.400	28,6
Occupazione e lavoro	31.150	40,7	4.634.195.667	32,6
Reti e servizi digitali	74	0,1	38.933.224	0,3
Ricerca e innovazione	526	0,7	82.368.694	0,6
Totale	**76.621**	**100,0**	**14.201.522.080**	**100,0**

Fonte: elaborazione IFEL-Area Studi e statistiche territoriali su dati OpenCoesione aggiornati al 31.12.2025

progetti FESR relativi a trasporti e mobilità pesano per l'11,2% del totale dei costi rendicontabili.

Dei 14,2 miliardi di euro FSE+ di costi rendicontabili si evidenzia che il 32,6% è associato a interventi che riguardano l'occupazione e il lavoro, il 28,6% a progetti in tema di istruzione/formazione e il 26,6% a operazioni per l'inclusione sociale e la salute.

Attraverso OpenCoesione si può risalire alla taglia finanziaria dei progetti: nel caso del FESR (tabella 4) i progetti fino a 150 mila euro sono il 69,9%, mentre nel caso del FSE+ (tabella 5) la percentuale sale al 90,4%. Per l'FSE+ si conferma, anche rispetto ai passati cicli di programmazione, una tendenza spiccata alla micro-progettualità dal momento che oltre il 37% degli interventi ha un costo rendicontabile inferiore ai 10 mila euro.

Rispetto ai grandi interventi con un costo rendicontabile superiore ai 5 milioni di euro, il FESR ne conta il l'1,7% del totale, mentre il FSE+ soltanto lo 0,3%.

Le differenze appena mostrate di fatto si spiegano considerando la diversa natura dei Fondi, il FESR che prevede interventi più di tipo infrastrutturale e il FSE+ che si caratterizza per una moltitudine di micro-progetti a supporto dei singoli cittadini.

Tabella 4. I progetti FESR 2021-2027, per classe di costo rendicontabile UE dell'intervento, 31 dicembre 2025

Classe di costo rendicontabile UE	N. progetti		Costo rendicontabile UE	
	v.a.	% sul totale	v.a. (euro)	% sul totale
Fino a 150.000 euro	27.336	69,9	1.148.580.085	5,2
Da 150.000 a 500.000 euro	7.178	18,4	1.981.979.165	8,9
Da 500.000 a 1.500.000 euro	2.534	6,5	2.201.660.452	9,9
Da 1.500.000 a 5.000.000 euro	1.371	3,5	3.678.593.419	16,5
Da 5.000.000 a 10.000.000 euro	361	0,9	2.558.874.256	11,5
Da 10.000.000 a 50.000.000 euro	244	0,6	4.880.181.489	21,9
Oltre 50.000.000 euro	77	0,2	5.823.631.622	26,1
Totale	**39.101**	**100,0**	**22.273.500.488**	**100,0**

Fonte: elaborazione IFEL-Area Studi e statistiche territoriali su dati OpenCoesione aggiornati al 31.12.2025

Tabella 5. I progetti FSE+ 2021-2027, per classe di costo rendicontabile UE dell'intervento, 31 dicembre 2025

Classe di costo rendicontabile UE	N. progetti		Costo rendicontabile UE	
	v.a.	% sul totale	v.a. (euro)	% sul totale
Fino a 10.000 euro	28.623	37,4	155.632.277	1,1
Da 10.000 a 30.000 euro	12.247	16,0	225.986.993	1,6
Da 30.000 a 50.000 euro	6.872	9,0	283.795.270	2,0
Da 50.000 a 100.000 euro	16.767	21,9	1.236.000.061	8,7
Da 100.000 a 150.000 euro	4.722	6,2	576.339.486	4,1
Da 150.000 a 500.000 euro	4.876	6,4	1.315.444.413	9,3
Da 500.000 a 1.000.000 euro	1.122	1,5	806.697.948	5,7
Da 1.000.000 a 5.000.000 euro	1.151	1,5	2.548.233.013	17,9
5.000.000 euro e oltre	241	0,3	7.053.392.618	49,7
Totale	**76.621**	**100,0**	**14.201.522.080**	**100,0**

Fonte: elaborazione IFEL-Area Studi e statistiche territoriali su dati OpenCoesione aggiornati al 31.12.2025

Come detto prima, si possono analizzare i progetti per tipologia di beneficiari dopo aver classificato i soggetti grazie alla *'join'* tra codici fiscali degli open data relativi alla politica di coesione e quelli relativi all'anagrafe degli enti della PA. Analizzando i 39 mila progetti FESR (figura 2), si nota come siano gli operatori privati e le imprese[10], con l'82,0% degli interventi a titolarità, i primi beneficiari FESR, seguiti a grande distanza dai comuni (10,5%).

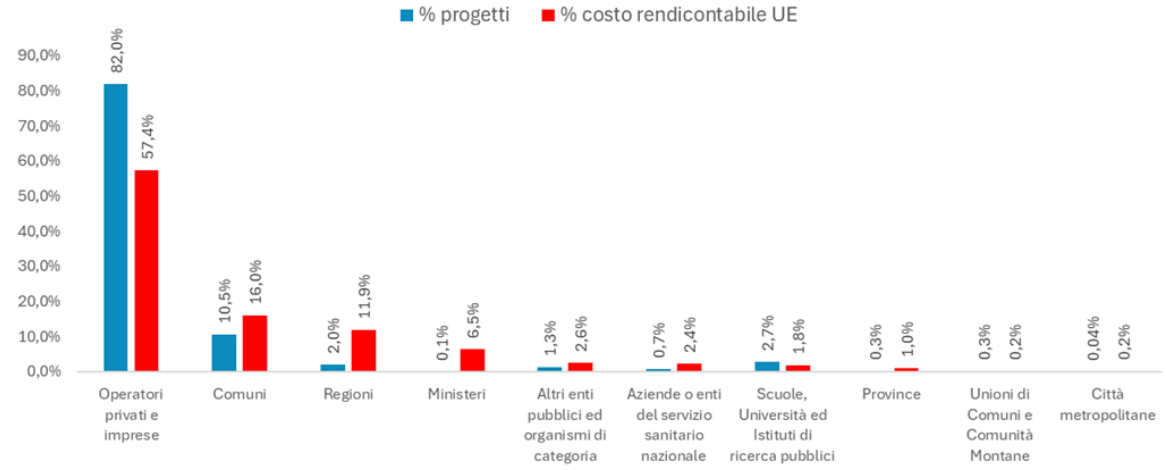
Considerando il valore complessivo del costo rendicontabile a valere sulle risorse FESR 2021-2027 (figura 2), si nota come dopo gli operatori privati (57,4%) la percentuale più rilevante sia quella dei comuni, che gestiscono risorse per un valore complessivamente superiore a 3,5 miliardi di euro, pari al 16,0% del costo totale. Di fatto i comuni sono i primi beneficiari pubblici del FESR. Seguono le amministrazioni regionali con l'11,9% delle risorse.

Si segnala che esiste una casistica ridotta di progetti gestiti da più beneficiari contemporaneamente (c.d. multibeneficiari) per i quali non è possibile scomporre le risorse associate al progetto tra singoli enti).

Anche nel FSE+ (figura 3) sono i privati i primi beneficiari in termini di numerosità dei progetti (oltre il 70% dei 76 mila interventi), seguiti questa volta da scuole, università ed istituti di ricerca

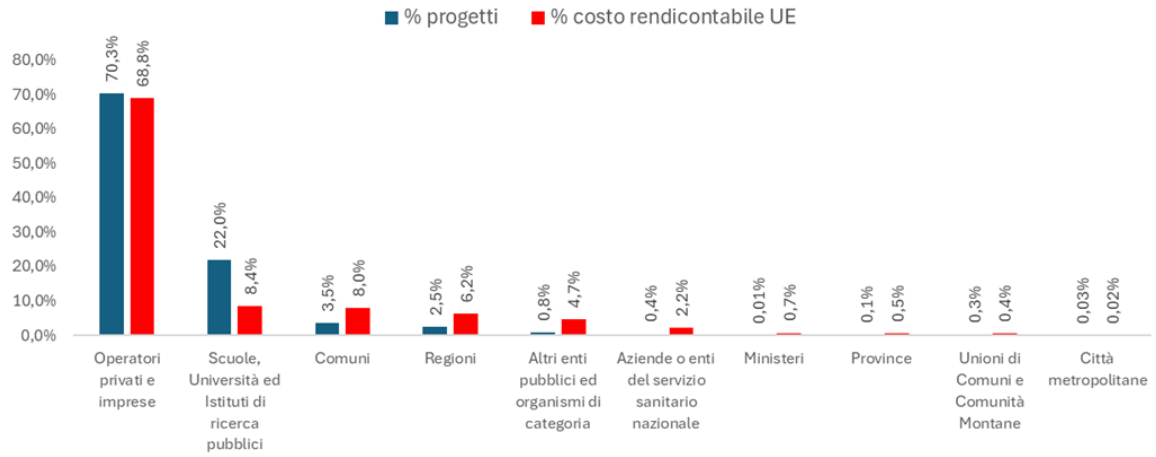
10 Includono tutti quei soggetti a capitale pubblico e/o pubblico-privato.

Figura 2. I progetti FESR 2021-2027, per tipologia di soggetto beneficiario, 31 dicembre 2025



Fonte: elaborazione IFEL-Area Studi e statistiche territoriali su dati OpenCoesione aggiornati al 31.12.2025

Figura 3. I progetti FSE+ 2021-2027, per tipologia di soggetto beneficiario, 31 dicembre 2025



Fonte: elaborazione IFEL-Area Studi e statistiche territoriali su dati OpenCoesione aggiornati al 31.12.2025

pubblici (22,0%). In termini di risorse, si confermano i privati come prima tipologia di soggetto beneficiario (il 68,8% dei costi è in capo a loro), seguiti dal comparto istruzione (8,4%) e dalle amministrazioni comunali (8,0%).

Rispetto ai dati di chiusura del ciclo 2014-2020, aggiornati al 31 dicembre 2023, al momento il coinvolgimento dei privati risulta più spiccato: nel settennio precedente i beneficiari privati gestivano il 44% dei costi rendicontabili UE FESR (contro l'attuale 57%) e il 24% nel caso del FSE (vs circa il 69% del FSE+ di dicembre 2025). Sarà interessante studiare l'evoluzione di tali percentuali con l'avanzamento della politica di coesione che entrerà nel vivo nei prossimi mesi/anni.

5. I comuni beneficiari di progetti FESR e FSE+ 2021-2027

Appare interessante, infine, approfondire il livello di coinvolgimento dei comuni italiani nell'utilizzo di risorse FESR e FSE+: alla data del 31 dicembre 2025, i comuni beneficiari dei 6.808 progetti ascrivibili ai due Fondi sono 2.427, il 30,7% del totale (tabella 6). In specifiche realtà regionali tale percentuale supera addirittura il 70% (Toscana e Calabria) e l'80% (Emilia-Romagna e Marche), mentre in altre è molto contenuta (Trentino-Alto Adige, Abruzzo, Molise e Basilicata).

Sul versante delle risorse, le regioni in cui si concentrano le quote maggiori di risorse a titolarità comunali sono la Campania (14,7%), l'Emilia-Romagna (10,9%) e la Lombardia (10,5%).

Tabella 6. I progetti FESR e FSE+ 2021-2027 con comuni beneficiari, per regione, 31 dicembre 2025

Regione	N. comuni italiani (a)	N. comuni beneficiari di progetti (b)	N. progetti con comuni beneficiari	% di comuni beneficiari di progetti (b/a)	Costo rendicontabile UE	
					v.a. (mln euro)	% sul totale
Piemonte	1.180	170	380	14,4	296	6,3
Valle d'Aosta	74	11	12	14,9	7	0,1
Lombardia	1.502	184	331	12,3	494	10,5
Trentino-Alto Adige	282	2	3	0,7	0,40	0,01
Veneto	560	165	277	29,5	141	3,0
Friuli-Venezia Giulia	215	60	174	27,9	57	1,2
Liguria	234	76	193	32,5	166	3,5
Emilia-Romagna	330	293	1.539	88,8	510	10,9
Toscana	273	197	1.019	72,2	411	8,8
Umbria	92	43	94	46,7	35	0,7
Marche	225	189	330	84,0	116	2,5
Lazio	378	133	406	35,2	326	6,9
Abruzzo	305	22	34	7,2	25	0,5
Molise	136	5	12	3,7	11	0,2
Campania	550	299	870	54,4	691	14,7
Puglia	257	117	261	45,5	345	7,3
Basilicata	131	7	15	5,3	25	0,5
Calabria	404	298	470	73,8	297	6,3
Sicilia	391	117	278	29,9	464	9,9
Sardegna	377	39	110	10,3	279	5,9
Totale	**7.896**	**2.427**	**6.808**	**30,7**	**4.698**	**100,0**

Fonte: elaborazione IFEL-Area Studi e statistiche territoriali su dati OpenCoesione, 2026

Tabella 7. I progetti FESR e FSE+ 2021-2027 con comuni beneficiari, per classe di ampiezza demografica, 31 dicembre 2025

Classe di ampiezza demografica	N. comuni italiani (a)	N. comuni beneficiari di progetti (b)	N. progetti con comuni beneficiari	% di comuni beneficiari di progetti (b/a)	Costo rendicontabile UE	
					v.a. (mln euro)	% sul totale
0 - 1.999 ab.	3.550	605	853	17,0	203	4,3
2.000 - 4.999 ab.	1.969	589	1.057	29,9	330	7,0
5.000 - 9.999 ab.	1.173	484	1.173	41,3	364	7,7
10.000 - 19.999 ab.	682	354	1.064	51,9	391	8,3
20.000 - 59.999 ab.	424	302	1.306	71,2	800	17,0
60.000 - 249.999 ab.	87	83	912	95,4	1.345	28,6
>= 250.000 ab.	11	11	443	100,0	1.265	26,9
Totale	**7.896**	**2.427**	**6.808**	**30,7**	**4.698**	**100,0**

Fonte: elaborazione IFEL-Area Studi e Statistiche Territoriali su dati OpenCoesione, 2026

Con riferimento alla classe di ampiezza demografica dei comuni, deducibile dall'incrocio tra codice Istat comunale del soggetto beneficiario, importato da BDAP, e dato Istat del bilancio demografico relativo alla popolazione residente, si nota un tasso di coinvolgimento degli enti direttamente proporzionale alla numerosità della popolazione: all'aumentare della taglia demografica cresce anche la percentuale di comuni beneficiari, che passano dal 17,0% della classe con meno di 2 mila abitanti, fino al 100% delle 11 grandi città italiane (tabella 7). Lo stesso si può dire delle risorse assegnate ai comuni, che vanno dal 4,3% del totale nei piccolissimi comuni a circa il 27/28% nelle città con più di 60 mila residenti.

Riflessioni conclusive

L'analisi di dettaglio esposta nei due paragrafi precedenti, sia su livello nazionale che comunale nel ciclo 2021-2027, fa emergere con nettezza come la ricchezza di informazioni che OpenCoesione ha saputo nel tempo rendere fruibile è un patrimonio oggi più unico che raro nella misurazione delle politiche pubbliche di investimento.

I risultati ottenuti negli ultimi anni dalla comunità scientifica e dalle esperienze di monitoraggio civico che hanno utilizzato OpenCoesione mostrano come la disponibilità di dati aperti possa produrre effetti che vanno oltre la trasparenza amministrativa. Le evidenze raccolte in letteratura documentano una crescente capacità degli open data di sostenere processi di accountability pubblica, partecipazione civica, analisi territoriale e valutazione delle politiche. La sfida per il prossimo ciclo di programmazione consiste nel consolidare questo patrimonio informativo, ampliandone l'interoperabilità con altre basi dati pubbliche e rafforzando il collegamento tra monitoraggio, valutazione e decisione pubblica. In questa prospettiva, OpenCoesione può evolvere da strumento di trasparenza a vera e propria infrastruttura nazionale per il policy learning della politica di coesione.

Appare evidente, inoltre, che, soprattutto in una fase di *mid-term review*, gli open data possono diventare una tecnologia istituzionale per rendere la coesione più adattiva e più orientata a impatto. Il passaggio chiave è trattarli come infrastruttura di policy learning: un sistema in cui dati aperti, standard, capacità analitiche e cicli decisionali sono integrati. Questo è sicuramente ancora uno dei principali punti di debolezza del policy designer in Italia, che tende ad orientare i propri processi decisionali nelle politiche pubbliche ricorrendo, sovente, più 'alla pancia' che 'alla testa'. Per questo motivo sarebbe auspicabile valutare l'effetto della trasparenza tempestiva su performance attuativa (ritardi, riallocazioni, qualità del portafoglio) e misurare l'uso effettivo degli open data nelle decisioni (non solo accessi): anche banalmente attraverso citazioni in atti, deliberazioni, riprogrammazioni, audit, relazioni annuali ecc. È una strada ancora in salita se si pensa al 'c'era una volta' con cui si è iniziato questo articolo, ma a distanza di quasi venti anni proviamo a vedere la 'botte mezza piena': sulla politica di coesione praticamente abbiamo una conoscenza più che esaustiva degli investimenti condotti con risorse straordinarie. Certo non vorremmo trascorre altri venti anni per collegare formalmente i dataset chiave (attuazione, milestone, output, target ecc.) a routine di revisione periodica, previste dalla governance di qualsiasi ciclo di politiche di investimento pubblico, affinché la trasparenza diventi parte della macchina decisionale.

Bibliografia

Albanese G., de Blasio G. (2018), La politica europea di coesione in Italia: evidenze empiriche e interpretazioni, in Pellegrini G., Tortorella W. (a cura di), *L'impatto della politica di coesione in Europa e in Italia - Documento di valutazione n. 11*, Roma, Senato della Repubblica, Ufficio Valutazione Impatto

Buttiglione P.L., Reggi L. (2015), Il monitoraggio civico delle politiche di coesione e lo sviluppo di comunità civiche, *Prisma: economia, società, lavoro,* VI, n.1, pp.81-97

Collesi P., De Luca S. (2022), *Understanding the territory through data, data journalism and civic engagement: the Asoc project from Italy to Europe*, European Conference on Quality in Official Statistics, Vilnius, 7-10 June 2022

Fusacchia I., Armenise M., Mantuano M., Tabasso M. (2024), Imprese e Politica di Coesione in Italia. Analisi e Prospettive dai Dati Integrati, *Rivista Economica del Mezzogiorno*, n.4, pp.689-710

Janssen M., Charalabidis Y., Zuiderwijk A. (2012), Benefits, Adoption Barriers and Myths of Open Data and Open Government, *Information Systems Management (ISM)*, 29, n.4, pp.258-268

Reggi L., Dawes S. (2016), Open Government Data Ecosystems: Linking Transparency for Innovation with Transparency for Participation and Accountability, in Scholl J.H., Glassey O., Janssen M., Klievink B., Lindgren I., Parycek P., Tambouris E., Wimmer A.M., Janowski T., Sá Soares D. (Eds.), *Electronic Government: 15th IFIP WG 8.5 International Conference, EGOV 2016, Guimarães, Portugal, September 5-8, 2016,* Cham, Springer International Publishing, pp.74-86

Giorgia Marinuzzi
giorgia.marinuzzi@fondazioneifel.it
Responsabile dell'Area Studi e statistiche territoriali di IFEL-Fondazione ANCI. Esperta di economia dello sviluppo e del territorio, cultore della materia in Statistica economica e membro del collegio docenti della Scuola di Dottorato in Scienze sociali ed economiche dell'Università di Roma La Sapienza. É autrice di numerosi articoli in materia di personale comunale, Fondi strutturali, PNRR, sociosanitario, gestioni associate comunali e Partenariato Pubblico Privato. Fra le pubblicazioni recenti si segnalano: con Romagno A. e Tortorella W., Lavorare in comune: cosa attrae e cosa spinge a lasciare, *Risorse Umane*, n.1/2026; con Coppola G., Destefanis S. e Tortorella W., The Effectiveness of European Structural and Investment Funds across European NUTS2 Regions. Institutional and Structural Determinants, *Scienze Regionali*, n.1/2026.

Walter Tortorella
walter.tortorella@fondazioneifel.it
Capo Dipartimento Economia locale-Direttore Scuola IFEL. Esperto di politiche pubbliche e sviluppo economico, è stato Direttore del Centro Documentazione e Studi dei Comuni italiani ANCI-IFEL, Direttore Ufficio Studi della Fondazione Cittalia ANCI Ricerche. È autore di numerosi articoli in materia di public management e politiche di sviluppo. Fra le pubblicazioni recenti si segnalano: con Romagno A. e Marinuzzi G., Lavorare in comune: cosa attrae e cosa spinge a lasciare, *Risorse Umane*, n.1/2026; con Coppola G., Destefanis S. e Marinuzzi G., The Effectiveness of European Structural and Investment Funds across European NUTS2 Regions. Institutional and Structural Determinants, *Scienze Regionali*, n.1/2026.

VisitINPS e l'uso dei dati amministrativi per la conoscenza e la valutazione delle politiche pubbliche

Maria De Paola
INPS, Università della Calabria

Monica Paiella
INPS, Università di Napoli Parthenope

Gianfranco Santoro
INPS

L'articolo analizza il ruolo dei dati amministrativi come risorsa strategica per la conoscenza e la valutazione delle politiche pubbliche, con particolare riferimento all'esperienza dell'Inps. Attraverso i programmi VisitINPS e VisitPNRR-Lavoro, l'Istituto promuove l'accesso sicuro ai microdati per la ricerca, sostenendo studi che contribuiscono alla progettazione e alla valutazione delle politiche di welfare e del lavoro. L'articolo discute vantaggi e limiti dei dati amministrativi e propone quattro linee di sviluppo per rafforzarne il potenziale.

VisitINPS and the use of administrative data for knowledge production and public policy evaluation

This article examines the role of administrative data as a strategic resource for evaluating public policies, focusing on the experience of the Italian National Social Security Institute (Inps). Through the VisitINPS and VisitPNRR-Lavoro programs, Inps provides secure access to administrative microdata for research, supporting studies that inform welfare and labour market policies. The article discusses the strengths and limitations of administrative data and outlines four priorities for future development.

DOI: 10.53223/Sinappsi_2026-02-5

Citazione
De Paola M., Paiella M., Santoro G. (2026), VisitINPS e l'uso dei dati amministrativi per la conoscenza e la valutazione delle politiche pubbliche, *Sinappsi*, XVI, n.2, pp.77-90

Parole chiave
Dati amministrativi
Integrazione dei dati
Valutazione delle politiche

Keywords
Administrative data
Data integration
Policy evaluation

Introduzione

Comprendere i fenomeni sociali ed economici richiede di osservare i comportamenti che li generano e di analizzare come questi evolvono nel tempo in risposta a shock economici, cambiamenti tecnologici e nuove politiche pubbliche. Allo stesso modo, per progettare politiche efficaci è necessario conoscere con precisione bisogni e destinatari, oltre a valutarne sistematicamente gli esiti. Non è sufficiente introdurre nuove misure, ma bisogna verificare se le misure raggiungono i target previsti, con quali risultati e per quanto tempo; se producono effetti inattesi; e come migliorare ciò che funziona o correggere ciò che non funziona. La valutazione è, quindi, parte integrante del ciclo di vita delle politiche.

In questo quadro, i dati rappresentano un'infrastruttura strategica che permette di quantificare la portata dei problemi da affrontare, di monitorare i processi in atto, di confrontare l'efficacia di policy alternative e, soprattutto, di identificare chi beneficia degli interventi e chi rischia di restarne escluso (Cole *et al.* 2020; Connelly *et al.* 2016; Einav e Levin 2013). Come evidenziato da Card *et al.* (2011), i dati amministrativi prodotti dalle amministrazioni pubbliche sono eccezionalmente ricchi; mettendo insieme informazioni derivanti da archivi fiscali, previdenziali, sanitari, scolastici e catastali è possibile ricostruire numerosi aspetti del comportamento socioeconomico degli individui lungo l'intero ciclo di vita, dall'istruzione al lavoro,

dal reddito alla salute e alla pensione. Nagaraj e Tranchero (2023), considerando l'esperienza del Census Bureau statunitense, che consente ai ricercatori di accedere ai dati riservati attraverso infrastrutture sicure dedicate, mostrano che l'utilizzo di questi dati è in costante crescita e che le pubblicazioni basate su dati amministrativi riservati tendono ad avere un impatto scientifico e politico maggiore rispetto ad altri studi. Nagaraj *et al.* (2020), analizzando il caso dei dati satellitari del programma Landsat della NASA, mostrano che una maggiore accessibilità ai dati favorisce soprattutto ricercatori provenienti da Paesi in via di sviluppo e istituzioni meno prestigiose, contribuendo a democratizzare la produzione scientifica. Inoltre, l'apertura dei dati incentiva nuove linee di ricerca e una maggiore varietà geografica e tematica degli studi.

In alcuni Paesi, tra cui Norvegia, Finlandia e Svezia, i dati amministrativi sono disponibili per la ricerca scientifica già da molti anni, rappresentando una componente consolidata delle infrastrutture informative per le scienze sociali. In altri contesti, tra cui il nostro Paese, ma anche realtà come il Regno Unito e gli Stati Uniti, la progressiva apertura dei dati amministrativi alla comunità scientifica rappresenta un cambiamento relativamente recente ma di grande rilevanza, che ha ampliato in modo significativo le possibilità di analisi empirica nelle scienze sociali.

Tra le diverse fonti di informazione, i dati amministrativi offrono vantaggi distintivi: copertura ampia della popolazione di riferimento, continuità temporale, aggiornamento frequente e possibilità di integrazione tra archivi. La digitalizzazione dei sistemi informativi pubblici ha moltiplicato le potenzialità di utilizzo a fini di ricerca e valutazione, affiancando e, per alcuni aspetti, superando per granularità e tempestività le indagini campionarie tradizionali.

Tra le basi informative più estese in Europa vi sono gli archivi amministrativi dell'Istituto nazionale della previdenza sociale (Inps) su lavoro, previdenza e welfare in Italia. Tali archivi, prodotti per finalità gestionali, costituiscono una miniera di informazioni fondamentali per analizzare le traiettorie lavorative individuali e il comportamento delle imprese, nonché per misurare l'efficacia delle politiche pubbliche e supportare le decisioni dei policymaker. La disponibilità di canali di accesso differenziati, standard elevati di sicurezza e procedure dedicate consentono oggi un uso sempre più ampio e qualificato di questi dati nel rispetto della riservatezza.

Questo contributo mostra come i dati amministrativi possano consentire analisi e valutazioni evidence-based, con particolare attenzione all'esperienza VisitINPS. L'articolo ha natura descrittivo-istituzionale e si propone di sistematizzare l'esperienza decennale del programma VisitINPS, offrendo una ricostruzione sistematica del percorso maturato nel corso degli anni. Dopo una sintesi dei vantaggi e delle criticità dei dati amministrativi rispetto ad altre fonti (paragrafo 1), il testo descrive la ricchezza informativa degli archivi Inps (paragrafo 2) e i canali attraverso cui l'Istituto rende disponibili i dati per la ricerca (paragrafo 3). Vengono quindi presentati il programma VisitINPS (paragrafo 4) e l'iniziativa VisitPNRR-Lavoro (paragrafo 5), evidenziandone obiettivi, meccanismi di accesso e risultati. Il contributo si chiude con alcune proposte operative volte ad ampliare l'accesso, favorire l'interoperabilità tra banche dati, integrare survey mirate e diffondere microdati, anche sintetici, a fini scientifici.

1. Dati amministrativi e dati di survey: potenzialità e limiti a confronto

Le principali fonti di dati per l'analisi socioeconomica

I dati utilizzati per analizzare i fenomeni socioeconomici e valutare l'efficacia delle politiche pubbliche provengono principalmente da tre fonti: dati amministrativi, indagini campionarie e fonti ibride.

I dati amministrativi sono prodotti da enti pubblici nell'ambito della gestione dei propri compiti istituzionali. Si tratta di registrazioni di eventi e prestazioni che hanno valore legale o amministrativo, come i flussi Uniemens trasmessi dai datori di lavoro all'Inps, il Casellario pensioni, i registri dell'Agenzia delle entrate o dell'Anpal, le anagrafi degli studenti e dei docenti. La loro finalità originaria non è la ricerca, ma la gestione efficiente dei servizi: proprio per questo, rappresentano una testimonianza diretta, continuativa e completa dei comportamenti individuali e collettivi.

Le indagini campionarie (survey), al contrario, vengono progettate con finalità di ricerca e sono basate su disegni statistici che garantiscono la rappresentatività della popolazione di riferimento. Tra le più note figurano le indagini di Istat, quelle di Inapp sul lavoro e sull'istruzione, le rilevazioni internazionali sul rendimento scolastico del Programme for International Student Assessment (PISA) e sulle competenze degli adulti del Programme for the International Assessment of Adult Competencies (PIAAC), e la Survey on Household Income and Wealth (SHIW) della Banca d'Italia. Le survey consentono di raccogliere informazioni che non è possibile ottenere dalle fonti amministrative, come le aspettative, le opinioni, le motivazioni individuali o la struttura patrimoniale delle famiglie.

Le fonti integrate o ibride, come i dati Invalsi o

AlmaLaurea, combinano invece i vantaggi dei due mondi, collegando registri amministrativi e dati rilevati tramite questionari. Queste basi offrono una visione più completa e multidimensionale dei fenomeni, ma richiedono un notevole sforzo di armonizzazione e governance.

Punti di forza e criticità dei dati amministrativi

Rispetto ai dati di survey, i dati amministrativi presentano diversi vantaggi strutturali che li rendono sempre più centrali per la ricerca applicata e la valutazione delle politiche.

Il principale punto di forza risiede nella copertura universale della popolazione di riferimento, e non solo di un campione. Ciò elimina gli errori campionari e consente analisi a livello territoriale con un elevato livello di dettaglio (comuni, province, distretti), nonché approfondimenti mirati su particolari segmenti della popolazione, come i lavoratori stagionali, i pensionati anticipati o i percettori di particolari misure di sostegno, che nelle survey risultano spesso sottorappresentati. Inoltre, i dati amministrativi vengono raccolti in modo continuativo e aggiornato per finalità gestionali. Ciò garantisce una tempestività e una granularità temporale (mensile, trimestrale) che le indagini campionarie, spesso con cadenza pluriennale, non possono offrire.

Un altro vantaggio riguarda l'accuratezza delle informazioni. Poiché derivano da atti e registrazioni ufficiali, i dati amministrativi non soffrono dei problemi di self-reporting tipici delle survey (errori di memoria, sottodichiarazioni di reddito, *bias* derivanti da fenomeni di desiderabilità sociale). Inoltre, non sono soggetti ai problemi di *non response*, che nelle indagini campionarie possono essere significativi e talvolta difficili da quantificare, compromettendo la rappresentatività anche quando si applicano pesi campionari. In generale, le informazioni sono oggettive e soggette a controlli automatici, aumentando l'affidabilità delle misure.

Gli archivi amministrativi permettono poi di seguire individui, imprese e famiglie nel tempo, ricostruendo traiettorie lavorative, previdenziali o demografiche. Questa dimensione longitudinale è raramente disponibile nelle indagini campionarie e costituisce un vantaggio cruciale per la valutazione delle politiche nel medio-lungo periodo. Infine, l'utilizzo dei dati già disponibili presso le amministrazioni pubbliche riduce la necessità di avviare nuove raccolte, generando significativi risparmi economici e diminuendo il *burden* amministrativo per cittadini e imprese.

Come evidenziato da Figlio *et al.* (2017), la possibilità di analizzare dati amministrativi riferiti all'intera popolazione apre opportunità di ricerca difficilmente raggiungibili attraverso indagini campionarie tradizionali. Oltre a garantire una maggiore potenza statistica e stime più precise, questi dati consentono di studiare eventi rari o situazioni specifiche, come ad esempio analisi basate sul confronto tra gemelli, effetti di eventi climatici estremi o chiusure di imprese. Permettono inoltre di analizzare gruppi normalmente poco rappresentati nelle indagini campionarie tradizionali, come gli individui agli estremi della distribuzione del reddito. La grande numerosità delle osservazioni consente, infine, di analizzare con maggiore dettaglio gli effetti eterogenei delle politiche pubbliche e delle pratiche educative su differenti gruppi della popolazione.

I vantaggi dei dati amministrativi si accompagnano tuttavia a limiti rilevanti, che emergono con chiarezza quando questi ultimi vengono confrontati con le survey.

Innanzitutto, i dati amministrativi vengono raccolti per finalità gestionali, non per la ricerca. Ciò comporta una limitata disponibilità di variabili di contesto (come, ad esempio, informazioni sul livello di istruzione dei genitori, sulle motivazioni o sulle competenze) e possibili lacune informative rispetto a concetti teorici di interesse analitico. Le categorie utilizzate nei registri possono poi non coincidere con quelle statistiche o economiche. Ad esempio, la definizione amministrativa di 'occupato' o 'disoccupato' può differire da quella adottata dall'Istat o da Eurostat, complicando i confronti.

Anche se raccolti in modo sistematico, i dati amministrativi possono incorporare errori, ritardi e disomogeneità dovuti ai processi gestionali. L'assenza di metadati esaustivi o documentazione sulle revisioni può compromettere la replicabilità delle analisi. Inoltre, ogni amministrazione adotta formati e standard propri. Collegare archivi diversi, come quelli dell'Inps, dell'Agenzia delle Entrate e del Ministero dell'Istruzione e del merito, richiede competenze tecniche, investimenti e, spesso, una base normativa che consenta il trattamento incrociato dei dati. La mancanza di identificativi univoci o la presenza di chiavi non condivise (es. codici SIDI rispetto a codice fiscale) complica ulteriormente i collegamenti.

La normativa sulla protezione dei dati personali (GDPR) costituisce un ulteriore vincolo sostanziale e impone restrizioni e procedure spesso complesse per l'accesso ai microdati. Per questo, i tempi di autorizzazione possono essere lunghi e disincentivare l'utilizzo da parte della comunità scientifica, soprattutto nei progetti con risorse o scadenze limitate.

Infine, rispetto alle survey, i dati amministrativi si caratterizzano per una minore flessibilità informativa. Le indagini campionarie permettono

di introdurre nuove domande e moduli in base alle esigenze di ricerca; i dati amministrativi, invece, riflettono la struttura dei processi gestionali, e modificare le variabili su cui vengono raccolte informazioni è molto più difficile.

Un bilancio comparativo

L'integrazione tra le due fonti rappresenta la via più efficace per lo sviluppo di sistemi informativi pubblici realmente orientati alla conoscenza e alla valutazione delle politiche. In quest'ottica, sarebbe auspicabile la creazione di ambienti sicuri per il *linkage* che, nel rispetto del GDPR, abilitino i collegamenti controllati tramite identificativi anonimizzati e procedure di *disclosure control*, ossia tecniche che impediscono la possibilità di risalire, direttamente o indirettamente, all'identità dei soggetti presenti nei dati, garantendo così la piena tutela della riservatezza. In quest'ultimo ambito, rientrano le soluzioni cosiddette privacy-preserving, nella forma di microdati sintetici[1], record linkage fiduciario[2] e tecniche di perturbazione o di computazione sicura.

In assenza di integrazione, i due mondi restano complementari ma separati: i registri amministrativi offrono copertura, tempestività e profondità longitudinale; le survey forniscono dimensioni concettuali non osservabili nei registri (aspettative, motivazioni, patrimonio, competenze). L'obiettivo dovrebbe essere quello di costruire, passo dopo passo, un ecosistema di interoperabilità che consenta di combinare queste potenzialità, mantenendo alte tutele per la privacy e la sicurezza dei dati e massimizzando il valore pubblico della conoscenza prodotta.

2. Inps come fonte conoscitiva

L'Inps rappresenta oggi una delle principali infrastrutture informative pubbliche in Europa nel campo del lavoro, della previdenza e del welfare. Accanto alla funzione storica di erogatore di prestazioni sociali, l'Istituto svolge un ruolo sempre più rilevante come produttore di evidenze utili per la ricerca e per la progettazione delle politiche pubbliche. Il suo patrimonio informativo, originato da attività gestionali e operative, consente di osservare in modo capillare le dinamiche del mercato del lavoro, l'evoluzione dei sistemi di protezione sociale e la distribuzione effettiva dei benefici pubblici.

Il perimetro operativo dell'Inps copre quasi tutta la popolazione attiva e pensionata italiana. Nel 2023 l'Istituto ha gestito dati relativi a circa 25 milioni di lavoratori (il 95% della forza lavoro totale), 1,7 milioni di imprese, 16 milioni di pensionati e milioni di beneficiari di prestazioni di sostegno al reddito, indennità di disoccupazione, congedi parentali, assegni familiari e bonus di vario tipo. A questi si aggiungono gli archivi contributivi e previdenziali, i flussi Uniemens inviati mensilmente dai datori di lavoro, le posizioni assicurative dei lavoratori autonomi e professionisti, le informazioni sul calcolo e sulla liquidazione delle pensioni, nonché le banche dati relative a prestazioni assistenziali e trasferimenti condizionati alla situazione economica delle famiglie. Questo insieme di archivi fornisce una rappresentazione pressoché completa della popolazione italiana sotto il profilo lavorativo, previdenziale e assistenziale, con un dettaglio temporale e territoriale difficilmente raggiungibile da altre fonti. Ciò rende l'Inps non solo un ente erogatore di servizi e prestazioni, ma anche un osservatorio strutturale sul funzionamento del welfare e del mercato del lavoro.

La progressiva valorizzazione del patrimonio informativo dell'Istituto ha mostrato come i dati amministrativi possano trasformarsi in uno strumento di analisi e di *policy evaluation*. L'uso dei microdati Inps consente, infatti, di: (1) individuare con precisione i beneficiari delle diverse misure di welfare, analizzando la distribuzione per età, genere, territorio e settore economico; (2) valutare l'efficacia degli interventi, verificando in che misura le politiche raggiungono i destinatari previsti e producono effetti su occupazione, reddito o condizioni familiari; (3) analizzare le traiettorie individuali e organizzative nel tempo, collegando percorsi lavorativi, previdenziali e assistenziali; (4) studiare le interazioni tra misure diverse, come il coordinamento tra ammortizzatori sociali, pensioni anticipate e sussidi di povertà; (5) costruire indicatori strutturali sulla sostenibilità del sistema e sull'equità distributiva delle prestazioni.

Grazie all'ampiezza e alla qualità delle informazioni amministrative raccolte, l'Inps costituisce una fonte informativa di rilievo per la ricerca e per la valutazione delle politiche pubbliche. I suoi archivi, estesi e aggiornati, possono supportare analisi empiriche e attività di programmazione da parte di amministrazioni e ricercatori. L'Istituto, attraverso la propria Direzione centrale Studi e ricerche, ha progressivamente sviluppato strumenti

1 Dati artificiali, generati per riprodurre la struttura statistica dei dataset reali senza contenere informazioni riconducibili a persone o imprese.

2 Il record linkage fiduciario consente di collegare archivi diversi tramite un ente terzo di fiducia, che genera chiavi di anonimizzazione comuni e impedisce che le amministrazioni o i ricercatori possano risalire all'identità dei soggetti. Si tratta di una soluzione per realizzare integrazioni sicure tra basi dati amministrative, nel pieno rispetto della normativa sulla protezione dei dati personali.

Tabella 1. Canali di accesso ai dati Inps

Canale di accesso	Destinatari principali	Livello di dettaglio	Modalità di accesso	Costi	Tempi medi	Vincoli normativi
Osservatori statistici	Ricercatori, giornalisti, professionisti, amministrazioni pubbliche, cittadini	Dati aggregati, indicatori e serie storiche	Consultazione online tramite portale Inps	Gratuiti	Immediati	Dati aggregati e anonimizzati; nessun accesso a microdati
Open Data	Comunità scientifica, media, società civile, cittadini	Dataset aggregati e anonimizzati scaricabili	Download diretto in formato aperto	Gratuiti	Immediati	Rispetto della normativa privacy; dati non identificativi
Fornitura dati a richiesta	Enti pubblici, enti di ricerca, organismi internazionali, soggetti privati autorizzati	Dataset personalizzati ed elaborazioni ad hoc	Richiesta formale all'Istituto	Gratuita per Enti appartenenti al SISTAN e a pagamento per altri (costi tecnici)	Variabili secondo complessità della richiesta	Accesso subordinato ad anonimizzazione (se richiesta da enti non SISTAN) e normativa sulla protezione dei dati
VisitINPS	Ricercatori universitari, centri di ricerca, istituzioni scientifiche selezionate tramite bando	Microdati amministrativi ad alta granularità	Accesso protetto tramite piattaforma sicura e postazioni dedicate	Generalmente gratuito per progetti selezionati	Dipendenti da bandi e autorizzazioni	Accesso controllato; impossibilità di esportare dati grezzi; disclosure control sugli output

Nota: le caratteristiche riportate si riferiscono alla disciplina vigente al momento della stesura del contributo.
Fonte: classificazione elaborata dagli Autori sulla base della documentazione istituzionale Inps

e procedure che consentono di bilanciare l'accesso ai dati con la tutela della riservatezza, promuovendo la diffusione della conoscenza e la valutazione basata sull'evidenza.

La duplice funzione gestionale e informativa riflette una trasformazione più ampia nella Pubblica amministrazione: i dati non sono più soltanto un prodotto collaterale dell'attività burocratica, ma un bene pubblico informativo che, se gestito con regole di sicurezza e governance adeguate, può alimentare analisi, valutazioni e decisioni più consapevoli.

In prospettiva, il patrimonio informativo Inps rappresenta una base imprescindibile per costruire un sistema integrato di analisi delle politiche del lavoro e del welfare. L'utilizzo di questi dati, in combinazione con altre fonti pubbliche – statistiche, fiscali, sanitarie, educative – può contribuire a delineare un modello di *policy intelligence* capace di migliorare la qualità della spesa sociale, monitorare l'impatto delle riforme e individuare con maggiore precisione i bisogni dei cittadini.

3. Inps e uso dei dati amministrativi ai fini della ricerca

Negli ultimi anni l'Inps ha sviluppato una strategia articolata per valorizzare il proprio patrimonio informativo e renderlo disponibile alla comunità scientifica e istituzionale, mantenendo al contempo elevati standard di sicurezza e tutela della privacy. Questa apertura si fonda su un principio chiave: i dati amministrativi, se gestiti in modo sicuro e controllato, possono generare conoscenza di interesse pubblico, migliorando la qualità delle politiche e la trasparenza dell'azione amministrativa.

L'accesso ai dati dell'Istituto si articola oggi attraverso quattro canali principali (tabella 1), differenziati per livello di approfondimento, destinatari e finalità: gli osservatori statistici, la sezione open data, la fornitura dati a richiesta e la piattaforma VisitINPS.

Gli osservatori statistici

Gli osservatori statistici rappresentano il primo livello di accesso, orientato alla diffusione sistematica e aggiornata di informazioni aggregate e alla comunicazione pubblica dei principali fenomeni del mercato del lavoro e del welfare. Ogni osservatorio – dedicato a temi come l'occupazione, la disoccupazione, le pensioni o le prestazioni di sostegno al reddito – fornisce indicatori aggiornati, serie storiche e database navigabili, disponibili in formato aperto e accompagnati da note metodologiche.

Accanto alla pubblicazione periodica di note statistiche che sintetizzano i principali trend e indicatori, gli osservatori mettono a disposizione database navigabili e scaricabili che consentono agli utenti di esplorare i dati secondo alcune combinazioni predefinite di variabili, tipicamente per regione, età, genere e territorio. Questo strumento permette a ricercatori, giornalisti, operatori pubblici e cittadini di accedere in modo autonomo a informazioni aggregate su occupazione, pensioni, prestazioni di sostegno al reddito e altre misure di welfare.

La sezione Open Data

Un secondo livello di accesso è costituito dalla sezione Open Data, dove vengono pubblicati dataset ed elaborazioni relativi a fenomeni di rilevanza pubblica, in forma aggregata e anonimizzata.

Questa iniziativa risponde ai principi di trasparenza e accessibilità delle informazioni della Pubblica amministrazione, promuovendo la diffusione della conoscenza e la partecipazione informata al dibattito pubblico. L'approccio open consente a una platea ampia, che va dalla comunità scientifica ai media, dalle organizzazioni della società civile ai singoli cittadini, di utilizzare i dati per analisi, verifiche o visualizzazioni personalizzate. Gli open data costituiscono dunque una porta di accesso democratica all'informazione pubblica, pur limitata, per ragioni di tutela della privacy, a forme di aggregazione che non consentono analisi micro.

La fornitura dati a richiesta

Un terzo canale è rappresentato dal servizio di fornitura dati a richiesta, destinato a soggetti pubblici o privati che necessitano di elaborazioni personalizzate non disponibili attraverso i canali standard. Su richiesta, l'Istituto può produrre dataset ad hoc costruiti in base a specifiche esigenze analitiche (periodo di riferimento, livello territoriale, variabili selezionate), nel rispetto della normativa sulla protezione dei dati personali e previo pagamento dei costi tecnici di preparazione, validazione e anonimizzazione. Questo canale risponde a esigenze di ricerca o di monitoraggio istituzionale più mirate, tipiche di amministrazioni centrali e regionali, enti di ricerca o organismi internazionali, che richiedono dati specifici per la valutazione di politiche pubbliche o per analisi comparative.

La piattaforma VisitINPS

Il quarto e più avanzato canale è rappresentato dal programma VisitINPS, istituito nel 2015 per consentire l'accesso sicuro ai microdati amministrativi dell'Inps per finalità di ricerca e valutazione delle politiche pubbliche. Attraverso bandi competitivi, il programma mette a disposizione di ricercatori e istituzioni scientifiche microdati ad alta granularità in un ambiente protetto, nel pieno rispetto della normativa sulla privacy.

I progetti sono selezionati sulla base della qualità scientifica, della fattibilità tecnica e della rilevanza per le aree di interesse dell'Istituto; l'accesso ai dati avviene esclusivamente in ambiente controllato, con verifica preventiva degli output (*disclosure control*).

VisitINPS rappresenta il livello più avanzato della strategia di accesso ai dati dell'Inps. Mentre osservatori statistici e open data privilegiano trasparenza e ampia diffusione delle informazioni, la fornitura a richiesta risponde a esigenze specifiche e VisitINPS rende possibile la ricerca scientifica sui microdati individuali in un ambiente sicuro.

Nel complesso, questa architettura multilivello consente di coniugare apertura dei dati, tutela della riservatezza e produzione di conoscenza utile alla valutazione delle politiche pubbliche.

4. Il programma VisitINPS

Lanciato nel 2015, VisitINPS è il principale programma istituzionale italiano dedicato a promuovere l'utilizzo dei dati amministrativi dell'Inps per la ricerca scientifica e la valutazione delle politiche pubbliche.

Il programma è coordinato dalla Direzione centrale Studi e ricerche (DCSR), che ne cura la gestione operativa, i rapporti con i ricercatori e la diffusione dei risultati. È finanziato attraverso sponsorizzazioni, erogazioni liberali e contributi della Ragioneria generale dello Stato (RGS) del Ministero dell'Economia e delle finanze (MEF), in particolare per i bandi dedicati a temi di genere.

La governance prevede un Responsabile scientifico, in carica biennale, affiancato da un Comitato scientifico composto da esperti accademici e dal Direttore della DCSR. Il Responsabile coordina le fasi del programma, supervisiona la valutazione dei progetti, presiede il Comitato della collana *WorkINPS Papers* e, su proposta del Comitato, può nominare *VisitINPS Senior Experts* incaricati di attività di supporto scientifico.

Un elemento distintivo di VisitINPS è la stretta connessione con il mondo universitario e della ricerca, reso possibile anche dalla presenza, all'interno della DCSR, di professori universitari collocati in aspettativa, che fungono da ponte tra comunità scientifica e amministrazione, contribuendo a garantire un dialogo continuo e una contaminazione virtuosa tra approcci accademici e obiettivi istituzionali.

L'accesso agli archivi Inps avviene attraverso bandi competitivi pubblici, aperti a ricercatori affiliati a università o enti di ricerca nazionali

e internazionali. Ogni progetto viene valutato in base a criteri di qualità scientifica, solidità metodologica, fattibilità tecnica e rilevanza per le politiche pubbliche. I progetti selezionati accedono ai microdati in un ambiente sicuro che garantisce la tutela della privacy. La piattaforma VisitINPS consente di lavorare direttamente su microdati anonimizzati, permettendo analisi longitudinali sofisticate. La disponibilità di dati longitudinali consente di analizzare carriere lavorative, valutare l'impatto delle riforme e studiare le dinamiche del welfare e delle disuguaglianze. Nel tempo il numero dei dataset disponibili è progressivamente aumentato, riflettendo sia l'evoluzione delle banche dati amministrative dell'Inps sia la volontà dell'Istituto di rendere accessibili nuove fonti informative alla comunità scientifica, ampliando le possibilità di ricerca.

L'infrastruttura tecnica comprende server dedicati, software per la gestione di grandi dataset e postazioni di lavoro presso sedi Inps o centri accreditati, così da assicurare un ambiente sicuro e controllato, conforme alle migliori pratiche internazionali (*Safe Rooms* e *Secure Data Environments*, si veda ad esempio UK Statistics Authority 2015). I dati non possono essere estratti e i risultati delle analisi vengono sottoposti a verifiche di *disclosure control*, per escludere ogni rischio di re-identificazione. Il modello, adottato anche in altri Paesi, si ispira al framework internazionale dei *Five Safes*, che disciplina la sicurezza del progetto, dei ricercatori, dei dati, degli ambienti di accesso e dei risultati diffusi (Desai *et al.* 2016; Ritchie 2017). Tale framework individua cinque dimensioni fondamentali del rischio: l'appropriatezza del progetto di ricerca (*Safe Projects*), l'affidabilità e la formazione dei ricercatori autorizzati (*Safe People*), la protezione e anonimizzazione dei dati (*Safe Data*), la sicurezza degli ambienti di accesso ai dati (*Safe Settings*) e il controllo dei risultati diffusi per evitare rischi di identificazione (*Safe Outputs*).

Un confronto più sistematico con alcune esperienze internazionali aiuta a collocare VisitINPS nel panorama delle infrastrutture sicure per l'accesso ai dati amministrativi. In particolare, il modello statunitense dei Federal Statistical Research Data Centers, coordinato dal Census Bureau, e quello britannico dei Secure Research Services/ Trusted Research Environments rappresentano due riferimenti consolidati[3]. Il primo è basato su una rete nazionale di centri fisici sicuri, attraverso i quali ricercatori autorizzati possono accedere a microdati riservati del sistema statistico federale, inclusi dati censuari, survey, dati *linked employer-employee* e archivi amministrativi. Il secondo si caratterizza per una maggiore articolazione istituzionale, con diversi ambienti sicuri gestiti da organismi quali l'Office for National Statistics (ONS), UK Data Service e ADR UK, accomunati dall'adozione del framework dei *Five Safes* e da procedure di accreditamento dei ricercatori, valutazione dei progetti e controllo degli output. Rispetto a tali esperienze, VisitINPS presenta una scala istituzionale più circoscritta, essendo centrato sul patrimonio informativo di un singolo grande ente di welfare, ma condivide con esse i principi fondamentali dell'accesso regolato, della selezione scientifica dei progetti, della non esportabilità dei dati grezzi e del controllo preventivo dei risultati diffusi.

Il confronto (tabella 2) mostra che VisitINPS si colloca pienamente nella traiettoria internazionale degli ambienti sicuri per la ricerca su microdati amministrativi, pur con alcune specificità. A differenza del modello statunitense, che opera come rete inter-agenzia del sistema statistico federale, VisitINPS è costruito attorno a un singolo ente titolare di archivi amministrativi di eccezionale ampiezza nel campo del lavoro e del welfare. Rispetto al modello britannico, più policentrico e orientato alla costruzione di Trusted Research Environments (TRE) trasversali a più domini amministrativi, VisitINPS appare maggiormente specializzato, ma anche più direttamente collegato alle esigenze conoscitive e valutative dell'istituzione che produce e gestisce i dati.

In dieci anni di attività, secondo la più recente sintesi istituzionale disponibile (Inps 2025), VisitINPS ha prodotto risultati significativi sia in termini quantitativi sia qualitativi[4]:

- ha raccolto circa 700 domande di partecipazione da parte di ricercatori e ricercatrici, dimostrando l'elevato interesse della comunità scientifica per l'accesso ai dati amministrativi Inps;
- ha assegnato 264 progetti di ricerca, coprendo una vasta gamma di temi relativi al mercato del lavoro, alle pensioni, al welfare e alla disuguaglianza;
- ha pubblicato 94 WorkINPS Papers, che diffondono i risultati delle ricerche e

3 Si vedano: *The Five Safes Framework*, versione 30 gennaio 2025 del UK Government Digital Service, disponibile all'indirizzo www.gov.uk/data-ethics-guidance/the-five-safes-framework e *Federal Statistical Research Data Centers*, versione 19 febbraio 2026 del U.S. Census Bureau, disponibile all'indirizzo www.census.gov/about/adrm/fsrdc.html (consultati il 28 maggio 2026).

4 La fonte citata riporta circa 700 domande, 264 progetti assegnati, 94 working paper e 8 convegni. La collana WorkINPS Papers risulta successivamente proseguita, con numerazione arrivata almeno al n. 121 nel maggio 2026.

Tabella 2. Confronto tra VisitINPS e alcune infrastrutture internazionali di accesso sicuro ai dati

	VisitINPS (Italia)	FSRDC/SRDC (Stati Uniti)	SRS/SecureLab/TRE (Regno Unito)
Governance	Programma coordinato dalla Direzione centrale Studi e ricerche dell'Inps, con Responsabile scientifico e Comitato scientifico.	Rete coordinata dal Census Bureau nell'ambito del sistema statistico federale e dell'Interagency Council on Statistical Policy.	Modello distribuito tra ONS Secure Research Service, UK Data Service SecureLab, ADR UK e altri Trusted Research Environments.
Modalità di accesso	Accesso tramite bandi competitivi annuali e postazioni protette; impossibilità di esportare dati grezzi; *disclosure control* sugli output.	Accesso a ricercatori approvati tramite sedi fisiche FSRDC; necessità di proposta scientifica, autorizzazione e procedure di sicurezza.	Accesso remoto o in ambienti sicuri a ricercatori accreditati; progetti valutati in base all'interesse pubblico e alla sicurezza del trattamento.
Ampiezza degli archivi	Archivi amministrativi Inps su lavoro, previdenza, welfare, imprese, contribuzioni, pensioni e prestazioni sociali.	Ampia gamma di dati federali riservati: censimenti, survey, dati amministrativi di vario tipo, inclusi *linked employer-employee* data.	Dati de-identificati, spesso collegati, provenienti da amministrazioni pubbliche, survey, archivi sanitari, educativi, fiscali e sociali.
Tempi medi di autorizzazione	Variabili in funzione del bando, della complessità del progetto e della predisposizione dei dati.	Possono essere lunghi: diversi mesi per le proposte standard e tempi maggiori in caso di coinvolgimento di più agenzie o data owner.	Variabili in base al TRE, al dataset e all'accreditamento richiesto; l'accesso è subordinato a verifiche su ricercatori, progetto e sicurezza.
Costi per il ricercatore	Generalmente gratuito per i progetti selezionati, salvo eventuali costi indiretti di partecipazione.	Non vi sono tariffe dirette del Census Bureau per il singolo ricercatore, ma possono esserci costi indiretti o fee istituzionali legate al centro di accesso ai dati.	Variabili a seconda del servizio, del dataset e dell'ente gestore; in alcuni casi sono previsti costi di accesso, preparazione dati o supporto.
Output prodotto	WorkINPS Papers, articoli scientifici, seminari, convegni annuali e contributi alla valutazione delle politiche pubbliche.	Working papers, articoli scientifici, analisi per finalità statistiche e di evidence-building, previa revisione degli output.	Articoli scientifici, report di policy, dataset documentati, evidenze per amministrazioni pubbliche e data owner.
Criticità documentate	Capacità amministrativa limitata, tempi di accesso, necessità di rafforzare documentazione e interoperabilità con altri archivi.	Iter autorizzativi complessi, tempi lunghi, costi indiretti e vincoli legati all'accesso fisico presso sedi dedicate.	Eterogeneità tra TRE, complessità delle procedure di accreditamento, costi e necessità di standard comuni tra infrastrutture.

Nota: Il confronto ha finalità descrittive e si riferisce alle principali caratteristiche dei sistemi al momento della stesura del contributo. Le modalità operative possono variare nel tempo e in funzione dello specifico dataset o programma di accesso.
Fonte: Elaborazione degli Autori sulla base della documentazione istituzionale delle infrastrutture considerate e della letteratura citata nel testo

contribuiscono al dibattito scientifico e pubblico;
- ha organizzato 8 convegni annuali di presentazione e discussione dei risultati;
- ha affiancato a queste attività una programmazione costante di seminari scientifici, con due incontri al mese, in cui i titolari dei progetti condividono risultati preliminari e avanzamenti delle proprie analisi.

Molti studi nati nell'ambito di VisitINPS sono stati successivamente pubblicati su riviste scientifiche internazionali, contribuendo alla diffusione di evidenze empiriche sul sistema italiano di welfare e sul funzionamento del mercato del lavoro. L'iniziativa ha, inoltre, favorito la formazione di giovani ricercatori sull'uso dei dati amministrativi, rafforzando competenze metodologiche e relazioni professionali.

La rilevanza scientifica delle ricerche sviluppate nell'ambito del programma è testimoniata anche dal fatto che numerosi lavori basati sui dati VisitINPS sono stati pubblicati sulle principali riviste internazionali di economia, tra cui *Econometrica, American Economic Review* e *The Review of Economic*

Studies (Bianchi *et al.* 2023; Fenizia 2022; Fenizia e Saggio 2024; Giupponi e Landais 2023; Sauvagnat e Schivardi 2024).

L'impatto del programma è stato oggetto di uno studio indipendente di Lepinteur e Nisticò (2025), pubblicato nella collana WorkINPS Papers, che ha analizzato longitudinalmente i percorsi dei ricercatori che hanno partecipato ai bandi per valutare l'impatto dell'accesso ai dati sulla loro produttività scientifica, visibilità accademica ed esiti occupazionali. Utilizzando modelli econometrici ad effetti fissi e strategie di Regression Discontinuity Design, lo studio ha mostrato che l'accesso ai dati amministrativi può influenzare significativamente le carriere accademiche[5]. I ricercatori selezionati hanno aumentato non solo la produttività complessiva, ma soprattutto la qualità delle proprie pubblicazioni, con una maggiore presenza su riviste di alto impatto, in particolare nell'economia del lavoro, ambito in cui i microdati Inps offrono un vantaggio comparativo decisivo. In parallelo, la pubblicazione dei risultati dei progetti VisitINPS nelle principali serie di working paper, inclusa la collana WorkINPS Papers, la maggiore partecipazione a conferenze internazionali e il rafforzamento delle reti professionali ne hanno accresciuto la visibilità scientifica. Nel complesso, questi elementi hanno migliorato la loro posizione nel mercato accademico, più per il potenziale di ricerca dimostrato e la qualità della ricerca svolta che per il semplice aumento del numero dei lavori prodotti. L'analisi evidenzia, tuttavia, alcune eterogeneità: i ricercatori provenienti da dottorati più prestigiosi traggono maggiori benefici in termini di carriera, e permangono asimmetrie di genere nei percorsi professionali, nonostante produzioni scientifiche di pari qualità. Inoltre, l'accesso ai dati ha ridotto la probabilità di ottenere posizioni nei dipartimenti di economia più prestigiosi a livello internazionale, ma ha aumentato le opportunità di inserimento in dipartimenti di alto livello nel sistema accademico italiano, contribuendo alla costruzione di carriere solide in ambito nazionale.

VisitINPS rappresenta dunque un modello istituzionale di successo nel panorama italiano ed europeo. Dimostra che è possibile coniugare rigore scientifico, tutela della privacy e utilità pubblica in un'iniziativa che alimenta la ricerca indipendente e, al tempo stesso, restituisce valore conoscitivo alle istituzioni. Il programma ha contribuito a creare una cultura dell'*evidence-based policy* all'interno dell'Inps e, più in generale, nella Pubblica amministrazione, rafforzando il legame tra produzione scientifica e decisione politica. In prospettiva, l'esperienza di VisitINPS offre un laboratorio di governance dei dati pubblici da cui trarre insegnamenti per altre amministrazioni: l'importanza di ambienti sicuri per la ricerca, procedure competitive e trasparenti di accesso, e la costruzione di una comunità scientifica aperta e interdisciplinare. In questo senso, VisitINPS non è soltanto un programma di ricerca, ma un'infrastruttura cognitiva che rende più informate, e potenzialmente più efficaci, le politiche pubbliche italiane.

5. Il programma VisitPNRR-Lavoro

Il programma sperimentale VisitPNRR-Lavoro nasce per favorire la produzione di analisi, studi ed evidenze scientifiche sulle politiche del lavoro e del welfare attuate nell'ambito del Piano nazionale di ripresa e resilienza (PNRR). L'iniziativa amplia e integra l'esperienza maturata con VisitINPS, proponendosi di sfruttare le potenzialità informative dei dati amministrativi per monitorare l'attuazione delle misure del Piano e valutarne l'impatto socioeconomico.

L'obiettivo principale del programma è rafforzare la capacità di valutazione delle politiche pubbliche connesse al PNRR, promuovendo la collaborazione tra amministrazioni, ricercatori e istituzioni accademiche. Attraverso l'utilizzo dei dati amministrativi già disponibili, in particolare quelli dell'Inps e della Ragioneria generale dello Stato, VisitPNRR-Lavoro mira a costruire un quadro informativo integrato che consenta di analizzare l'efficacia e la distribuzione territoriale degli investimenti e degli interventi previsti dal Piano in materia di lavoro, inclusione e coesione sociale. Le analisi promosse intendono rispondere a una serie di interrogativi cruciali, tra cui, a titolo di esempio, quali categorie di lavoratori e quali territori traggono maggior beneficio dagli interventi finanziati dal Piano; quali misure generano effetti occupazionali più stabili o duraturi, e in che misura le politiche del PNRR incidono sulle disuguaglianze di genere, generazionali o territoriali. In questo senso, VisitPNRR-Lavoro costituisce un laboratorio di valutazione basata sui dati, finalizzato a fornire evidenze utili alle istituzioni responsabili dell'attuazione del Piano e, più in generale, alla governance del sistema di politiche del lavoro.

Il programma è gestito dalla DCSR dell'Inps in collaborazione con la RGS. L'accesso ai bandi è riservato a professori, ricercatori universitari ed esperti di materie statistiche ed economiche

5 Per isolare gli effetti causali, Lepinteur e Nisticò (2025) utilizzano modelli a effetti fissi bidirezionali (Two-Way Fixed Effects) e strategie di Regression Discontinuity Designs che sfruttano la discontinuità nella probabilità di selezione in prossimità della soglia di punteggio delle proposte progettuali.

appartenenti a università, enti di ricerca, amministrazioni pubbliche o fondazioni, italiani e stranieri. I progetti selezionati ottengono l'accesso ai microdati Inps necessari alla ricerca, disponibili tramite la piattaforma VisitINPS, e ai dati RGS relativi ai progetti finanziati dal PNRR, che includono informazioni sulla localizzazione territoriale, sui soggetti coinvolti e sullo stato di avanzamento fisico, finanziario e procedurale delle misure[6]. L'accesso ai dati avviene sempre nel rispetto della normativa in materia di protezione dei dati personali, attraverso la piattaforma VisitINPS.

Un Comitato di indirizzo sovrintende all'iniziativa, individuando le aree tematiche prioritarie, selezionando i microdati di interesse e monitorando lo stato di avanzamento dei progetti. Il Comitato promuove inoltre la diffusione pubblica dei risultati, favorendo la pubblicazione dei lavori nella collana WorkINPS Papers e incoraggiando il dialogo tra ricercatori, policymaker e operatori istituzionali.

VisitPNRR-Lavoro rappresenta un passo ulteriore verso un modello di collaborazione strutturata tra ricerca e amministrazione, in cui i dati diventano strumento di trasparenza, accountability e apprendimento istituzionale. L'iniziativa contribuisce a rafforzare la capacità analitica del settore pubblico, rendendo possibile la valutazione sistematica degli interventi del PNRR in termini di efficacia, equità e impatto territoriale. Sebbene ancora in fase iniziale, essa offre un modello replicabile per altre aree di policy, dalla sanità all'istruzione, dalle politiche industriali a quelle ambientali, dimostrando che la messa a sistema delle informazioni amministrative può trasformarsi in una leva per il miglioramento continuo delle politiche pubbliche.

Le aree tematiche individuate nei primi due bandi di selezione riguardano: l'analisi degli effetti delle riforme e degli investimenti del PNRR sul mercato del lavoro e sul sistema di welfare; le valutazioni ex-ante delle riforme e/o degli investimenti e la definizione di metodologie per costruire indicatori utili alle verifiche ex-post; lo sviluppo di metodologie per la creazione, nella piattaforma informatica VisitINPS Scholars denominata VisitPNRR-Lavoro, di nuovi dataset anonimizzati a partire da microdati di interesse per la valutazione del PNRR e non di fonte Inps o RGS. Particolare attenzione è riservata agli obiettivi trasversali di parità di genere, riduzione delle disuguaglianze generazionali e territoriali, nonché alla transizione verde e digitale. Alla fine del 2025, a seguito della pubblicazione dei primi due bandi di selezione, sono stati avviati 19 progetti di ricerca che spaziano dalla valutazione delle politiche attive del lavoro ai processi di transizione digitale.

In prospettiva, la sfida sarà consolidare questa esperienza, garantendo la sostenibilità istituzionale del programma anche oltre la durata del PNRR e integrandolo stabilmente nelle strategie di valutazione e monitoraggio delle politiche del lavoro e del welfare.

Considerazioni conclusive

L'esperienza di VisitINPS rappresenta uno dei casi più significativi di valorizzazione dei dati amministrativi pubblici a fini di ricerca e valutazione in Italia. Il programma ha dimostrato che è possibile conciliare apertura e sicurezza, favorendo l'utilizzo dei dati da parte della comunità scientifica senza compromettere la tutela della privacy. Al tempo stesso, ha mostrato come la produzione di evidenze empiriche possa alimentare un circolo virtuoso tra conoscenza, ricerca e decisione pubblica: la scienza dei dati diventa così parte integrante della governance del welfare e del mercato del lavoro.

Tuttavia, la gestione di un programma complesso come VisitINPS ha fatto emergere alcune criticità strutturali che, se affrontate, potrebbero potenziarne ulteriormente il ruolo strategico. Una prima criticità riguarda la capacità amministrativa e tecnica necessaria a sostenere l'intero ciclo di vita dei progetti: selezione delle proposte, predisposizione dei dataset, assistenza ai ricercatori, gestione degli accessi, verifica degli output e diffusione dei risultati. L'elevata domanda di accesso ai dati e la crescente complessità delle richieste informative possono determinare tempi non trascurabili tra l'assegnazione del progetto e l'effettiva disponibilità dei dati. Tali tempi dipendono dalla natura del progetto, dal numero di archivi coinvolti, dalla necessità di costruire dataset personalizzati e dalle verifiche richieste per garantire la conformità alla normativa sulla protezione dei dati personali. Per questa ragione, il rafforzamento dell'organico dedicato al programma, in termini di competenze statistiche, informatiche, giuridiche e amministrative, rappresenta una condizione importante per migliorare la tempestività dell'accesso e la qualità del supporto offerto ai gruppi di ricerca.

Una seconda criticità riguarda la documentazione e la standardizzazione delle informazioni sui dataset. La disponibilità di metadati completi, dizionari delle variabili, note metodologiche, esempi di codice, informazioni sul *versioning* degli archivi e sulla

6 Si veda: *Dati PNRR disponibili per finalità di analisi e valutazione. Versione 31 luglio 2025* della Ragioneria generale dello Stato, disponibile all'indirizzo https://www.rgs.mef.gov.it/_Documenti/VERSIONE-I/attivita_istituzionali/monitoraggio/analisi_e_valutazione_del_pnrr/DatiPNRRdisponibili.pdf (consultato 28 maggio 2026).

tracciabilità delle modifiche normative ridurrebbe i tempi di apprendimento, migliorerebbe la replicabilità delle analisi e diminuirebbe il rischio di errori interpretativi. La complessità dei dati amministrativi, infatti, non risiede soltanto nella loro dimensione quantitativa, ma anche nella loro natura istituzionale: le variabili riflettono regole amministrative, procedure gestionali e cambiamenti normativi che devono essere correttamente compresi per produrre evidenze robuste.

Queste due criticità si traducono, per i ricercatori, in costi indiretti legati al tempo necessario per comprendere la struttura degli archivi, familiarizzare con le procedure di accesso, predisporre codici replicabili in ambienti protetti e interagire con il personale tecnico per la validazione degli output. Tali costi non sono necessariamente monetari, poiché l'accesso ai dati nell'ambito dei progetti selezionati è generalmente gratuito, ma possono incidere in modo significativo sulla fattibilità dei progetti, soprattutto per gruppi con risorse limitate. Le barriere all'ingresso risultano particolarmente rilevanti per i giovani studiosi, per i ricercatori non precedentemente esposti ai dati amministrativi Inps e per i gruppi di ricerca privi di reti consolidate con studiosi che abbiano già utilizzato la piattaforma. In questo senso, il rafforzamento dell'organico e il miglioramento della documentazione contribuirebbero anche a ridurre tali costi indiretti, rendendo l'accesso più inclusivo e meno dipendente dall'esperienza pregressa o dal capitale relazionale dei singoli gruppi di ricerca.

Infine, il pieno sfruttamento del potenziale di VisitINPS richiede un investimento sistematico nella formazione metodologica. Seminari, tutorial, moduli introduttivi sui principali archivi, linee guida per la costruzione dei dataset e attività di mentoring rivolte in particolare a dottorandi e giovani ricercatori potrebbero ridurre ulteriormente le barriere all'ingresso e ampliare la platea di utilizzatori qualificati. Ciò contribuirebbe a rendere il programma più inclusivo, a migliorare la qualità scientifica dei progetti e a favorire un uso più efficiente e responsabile dei dati amministrativi.

Alla luce di queste criticità, il rafforzamento della funzione dell'Inps come infrastruttura informativa per la ricerca e la valutazione delle politiche pubbliche richiede una strategia di sviluppo che combini ampliamento dell'accesso, interoperabilità tra archivi, integrazione con fonti campionarie e consolidamento degli strumenti di supporto alla ricerca.

Per rendere tali linee di sviluppo più operative, la tabella 3 ne sintetizza possibili modalità di attuazione, distinguendo tra orizzonte temporale, attori istituzionali da coinvolgere, sperimentazioni già avviate o ipotizzabili e benchmark internazionali di riferimento. La scansione proposta non ha carattere prescrittivo, ma individua alcune condizioni organizzative e istituzionali necessarie per trasformare le indicazioni generali in interventi progressivi e verificabili.

Questa operativizzazione evidenzia che le quattro linee di sviluppo non richiedono tutte lo stesso orizzonte temporale né lo stesso livello di coordinamento istituzionale. Alcune azioni, come il rafforzamento della documentazione, la predisposizione di tutorial e l'ampliamento delle attività formative, possono essere avviate nel breve periodo all'interno della governance esistente di VisitINPS. Altre, come l'interoperabilità tra archivi amministrativi o il collegamento sistematico tra dati amministrativi e survey, richiedono invece accordi interistituzionali, protocolli tecnici condivisi, valutazioni giuridiche specifiche e investimenti infrastrutturali. In questa prospettiva, VisitPNRR-Lavoro costituisce un primo laboratorio utile, perché sperimenta forme di integrazione tra dati amministrativi e informazioni progettuali relative al PNRR, offrendo indicazioni metodologiche e organizzative potenzialmente estendibili ad altri ambiti di policy.

Sul piano dell'accesso, l'obiettivo non è soltanto aumentare il numero dei progetti selezionati, ma anche diversificare i profili dei ricercatori coinvolti, includendo dottorandi, giovani studiosi e gruppi interdisciplinari di istituzioni italiane ed estere. Ciò consentirebbe di ampliare gli approcci metodologici e i campi di indagine, producendo evidenze empiriche più complete e utili alla comprensione delle dinamiche del mercato del lavoro e del welfare.

Sul piano dell'interoperabilità, il collegamento dei dati Inps con archivi fiscali, sanitari, educativi e territoriali aprirebbe la strada ad analisi multidimensionali e longitudinali in grado di osservare simultaneamente condizioni economiche, percorsi occupazionali, salute, istruzione e contesti locali. Per raggiungere questo obiettivo servono protocolli tecnici standardizzati, procedure amministrative condivise per l'accesso e infrastrutture sicure conformi alle norme sulla protezione dei dati personali e alla sicurezza informatica. Un'interoperabilità ben progettata trasformerebbe i dati amministrativi in uno strumento integrato di *policy evaluation*, capace di rispondere a domande sociali ed economiche complesse e di accrescere il valore informativo dell'intero sistema pubblico.

Sul piano dell'integrazione con fonti campionarie, le survey consentono di rilevare dimensioni che i registri amministrativi normalmente non osservano, quali composizione e ricchezza familiare, competenze, aspettative, motivazioni e

Tabella 3. Operativizzazione delle linee di sviluppo per il rafforzamento dell'accesso ai dati amministrativi

Linea di sviluppo	Orizzonte temporale realistico	Attori istituzionali da coinvolgere	Sperimentazioni o pilot	Benchmark internazionali
Ampliamento dell'accesso ai dati	**Breve-medio periodo**: ampliamento progressivo dei bandi, maggiore frequenza delle call, rafforzamento delle procedure di *onboarding*.	Inps-Direzione centrale Studi e ricerche e Inps Data Warehouse, Comitato scientifico VisitINPS e eventuali amministrazioni cofinanziatrici.	Estensione di seminari introduttivi, percorsi di *onboarding* per nuovi assegnatari, eventuali call tematiche per giovani ricercatori o gruppi interdisciplinari.	Federal Statistical Research Data Centers statunitensi; UK Secure Research Service; modelli nordici di accesso regolato ai registri amministrativi.
Interoperabilità tra archivi amministrativi	**Medio-lungo periodo**: sviluppo graduale, data la complessità tecnica, giuridica e organizzativa del *linkage* tra amministrazioni diverse.	Inps, Istat, Ministero dell'Economia e delle finanze/RGS, Agenzia delle Entrate, Ministero del Lavoro, Ministero dell'Istruzione e del merito, Ministero dell'Università e della ricerca, Ministero della Salute, Garante per la protezione dei dati personali.	Prima sperimentazione di integrazione tra dati Inps e dati RGS sul PNRR nell'ambito del programma VisitPNRR-Lavoro. In esame anche sperimentazioni di linkage con archivi fiscali, educativi e sanitari per specifiche analisi di policy.	Administrative Data Research UK; infrastrutture nordiche basate su registri integrati; Federal Statistical Research Data Centers statunitensi per l'accesso a dati linked.
Integrazione con survey mirate	**Medio periodo**: avvio tramite progetti pilota su popolazioni o misure specifiche, prima di una eventuale estensione sistematica.	Inps, Istat, Inapp, Banca d'Italia, Confindustria, università, enti di ricerca, Garante per la protezione dei dati personali.	Survey pilota collegate, in forma sicura e anonimizzata, a specifiche popolazioni osservate nei registri Inps; moduli sperimentali su aspettative, competenze, ricchezza familiare, percezione dei servizi e risposte comportamentali alle politiche.	Esperienze di survey-register linkage nei Paesi nordici; linked survey-administrative data presso UK Data Service/ADR UK; esperienze statunitensi di integrazione tra survey e administrative records.
Documentazione, supporto e formazione metodologica	**Breve periodo** per interventi di base; **medio periodo** per una infrastruttura stabile di formazione e metadatazione.	Inps-Direzione centrale Studi e ricerche e Inps Data Warehouse, Comitato scientifico VisitINPS, università, assegnatari dei progetti, esperti di *data management* e *disclosure control*.	Predisposizione di dizionari delle variabili, guide metodologiche, esempi di codice, tutorial, seminari periodici, moduli formativi per dottorandi e giovani ricercatori, repository controllati di documentazione.	Five Safes framework; UK Safe Researcher Training; pratiche di metadatazione e documentazione adottate da ONS Secure Research Service, UK Data Service SecureLab e FSRDC statunitensi.

Fonte: elaborazione degli Autori

percezioni. Progetti sperimentali di *survey linkage* permetterebbero di combinare la profondità concettuale delle indagini con la copertura e la dimensione longitudinale dei dati amministrativi, nel pieno rispetto delle tutele sulla privacy.

Infine, documentazione, supporto tecnico e formazione dovrebbero essere considerati parte integrante dell'infrastruttura di accesso. Metadati completi, dizionari delle variabili, guide metodologiche, esempi di codice, portali interattivi di consultazione e percorsi formativi strutturati non rappresentano semplici strumenti accessori, ma condizioni essenziali per rendere i dati realmente fruibili. Un ecosistema informativo ben documentato

e accompagnato da adeguati meccanismi di formazione riduce i costi di apprendimento, migliora la replicabilità degli studi, amplia la platea di utilizzatori qualificati e innalza la qualità complessiva delle analisi.

Le esperienze di VisitINPS e VisitPNRR-Lavoro mostrano che i dati amministrativi non sono solo un supporto tecnico alla gestione, ma una risorsa strategica per il policymaking. Mettere a sistema queste esperienze significa costruire un'infrastruttura nazionale della conoscenza pubblica, fondata su tre pilastri: accesso sicuro e regolato ai microdati, interoperabilità tra archivi amministrativi, collaborazione strutturata tra amministrazioni e comunità scientifica.

In un contesto in cui la valutazione delle politiche pubbliche si basa sempre più sull'integrazione tra dati, competenze e infrastrutture di ricerca, l'esperienza di VisitINPS mostra le potenzialità dei dati amministrativi come risorsa per la conoscenza e il supporto alle decisioni pubbliche. Consolidare e sviluppare questa esperienza significa rafforzare la capacità delle amministrazioni di produrre analisi rigorose, favorire la collaborazione con la comunità scientifica e contribuire a politiche più informate ed efficaci.

Bibliografia

Bianchi N., Bovini G., Li J., Paradisi M., Powell M. (2023), Career spillovers in internal labour markets, *The Review of Economic Studies*, 90, n.4, pp.1800-1831

Card D., Chetty R., Feldstein M.S., Saez E. (2011), Expanding Access to Administrative Data for Research in the United States, in *Ten Years and Beyond: Economists Answer NSF's Call for Long-Term Research Agendas*, Nashville (TN), American Economic Association

Cole S., Dhaliwal I., Sautmann A., Vilhuber L. (eds.) (2020), *Handbook on Using Administrative Data for Research and Evidence-based Policy*, Cambridge (MA), Abdul Latif Jameel Poverty Action Lab

Connelly R., Playford C.J., Gayle V., Dibben C. (2016), The role of administrative data in the big data revolution in social science research, *Social Science Research*, n.59, pp.1-12

Desai T., Ritchie F., Welpton R. (2016), *Five Safes: Designing Data Access for Research*, Economics Working Paper Series n.1601, Bristol, University of the West of England

Einav L., Levin J.D. (2013), *The Data Revolution and Economic Analysis*, NBER Working Paper n.19035, Cambridge (MA), National Bureau of Economic Research

Fenizia A. (2022), Managers and productivity in the public sector, *Econometrica*, 90, n.3, pp.1063-1084

Fenizia A., Saggio R. (2024), Organized crime and economic growth: Evidence from municipalities infiltrated by the mafia, *American Economic Review*, 114, n.7, pp.2171-2200

Figlio D., Karbownik K., Salvanes K.G. (2017), The promise of administrative data in education research, *Education Finance and Policy*, 12, n.2, pp.129-136

Giupponi G., Landais C. (2023), Subsidizing labour hoarding in recessions: The employment and welfare effects of short-time work, *The Review of Economic Studies*, 90, n.4, pp.1963-2005

Inps (2025), *XXIV Rapporto Annuale. Allegato VisitINPS*, Roma, Inps

Lepinteur A., Nisticò R. (2025), *"Based on Admin Data!". How Administrative Data Fosters Young Economists' Career*, WorkINPS paper n.93, Roma, Inps

Nagaraj A., Shears E., de Vaan M. (2020), Improving data access democratizes and diversifies science, *Proceedings of the National Academy of Sciences*, 117, n.38, pp.23490-23498

Nagaraj A., Tranchero M. (2023), *How does data access shape science? The impact of federal statistical research data centers on economics research*, NBER Working Paper n. 31372, Cambridge (MA), National Bureau of Economic Research

Ritchie F. (2017), The "Five Safes": a framework for planning, designing and evaluating data access solutions, *International Journal of Population Data Science*, 1, n.1 <DOI: 10.5281/zenodo.897821>

Sauvagnat J., Schivardi F. (2024), Are executives in short supply? Evidence from death events, *Review of Economic Studies*, 91, n.1, pp.519-559

UK Statistics Authority (2015), *Review of the UK Safe Researcher Training Scheme*, London, Statistics Authority

Maria De Paola
maria.depaola@inps.it
Dirigente di ricerca presso la Direzione centrale Studi e ricerche dell'Inps e Professore ordinario di Economia politica (in aspettativa) presso l'Università della Calabria. Presidente dell'Associazione italiana di economia del lavoro. I suoi interessi di ricerca riguardano principalmente l'economia del lavoro, l'economia dell'istruzione, l'economia di genere, la valutazione delle politiche pubbliche e l'economia sperimentale. Ha inoltre ricoperto ruoli di responsabilità istituzionale presso l'Università della Calabria, tra cui quello di Direttore di Dipartimento e Pro-rettore.

Monica Paiella
monicapiacecilia.paiella@inps.it
Dirigente di ricerca presso la Direzione centrale Studi e ricerche dell'Inps e Professore ordinario di Economia politica (in aspettativa) presso l'Università di Napoli Parthenope. Membro dello Steering Committee del CEPR Research Policy Network on Household Finance e CEPR Network Fellow. Fra le pubblicazioni recenti si segnalano: The Anatomy of the Wealth Effect, *Review of Economics and Statistics*, 2017 e Employment Growth, Inflation, and the Distribution of Household Labour Income, *Journal of Economic Inequality*, forthcoming.

Gianfranco Santoro
gianfranco.santoro@inps.it
Direttore della Direzione centrale Studi e ricerche dell'Inps dal 2023. In precedenza ha ricoperto il ruolo di coordinatore generale statistico attuariale dello stesso Istituto. È stato componente del Nucleo di coordinamento della politica economica presso la Presidenza del Consiglio dei Ministri tra il 2016 e il 2018.

Istruzione, salari ed eterogeneità delle imprese: evidenze da dati employer-employee linked

Irene Brunetti
INAPP

Michelangelo Filippi
Laboratorio Revelli

Valentina Gualtieri
INAPP

Andrea Ricci
INAPP

Lo studio analizza come le politiche retributive delle imprese contribuiscano ai rendimenti salariali dell'istruzione, considerando eterogeneità delle imprese. L'analisi utilizza un dataset integrato che combina Uniemens (Inps), Comunicazioni Obbligatorie (MLPS) e dati RIL Inapp, applicando una strategia econometrica a due stadi basata su modelli AKM e regressioni ad effetti fissi. I risultati mostrano che le politiche salariali influenzano positivamente, seppur debolmente, i premi dei laureati; che i differenziali retributivi riflettono processi di sorting non uniformi lungo la distribuzione del capitale umano; e che le imprese trasferiscono solo parzialmente e nel medio periodo i guadagni di produttività ai lavoratori più qualificati.

Education, wages, and firm heterogeneity: evidence from linked employer-employee data

The study examines how firms' wage setting practices shape the wage returns to education, accounting for firm's heterogeneity. Using a unique dataset combining Uniemens (Inps), COB (MLPS) with RIL-Inapp firm data, a two stage AKM fixed effects strategy shows that firm pay policies weakly raise graduates' wage premiums; sorting patterns differ across human capital levels; and productivity gains are only partly and slowly passed on to highly educated workers.

DOI: 10.53223/Sinappsi_2026-02-6

Citazione
Brunetti I., Filippi M., Gualtieri V., Ricci A. (2026), Istruzione, salari ed eterogeneità delle imprese: evidenze da dati employer-employee linked, *Sinappsi*, XVI, n.2, pp.91-105

Parole chiave
Employer-employee data
Imprese
Politiche salariali

Keywords
Employer-employee data
Enterprises
Wage setting

Introduzione

La persistente stagnazione delle retribuzioni reali rappresenta uno dei principali temi del dibattito economico e istituzionale sul mercato del lavoro italiano (OECD 2024; Inps 2024; Banca d'Italia 2025). Parallelamente, il crescente interesse della letteratura per le determinanti della dispersione salariale ha riportato al centro dell'analisi il ruolo delle imprese nella formazione dei differenziali retributivi. Numerosi studi evidenziano come la persistente debolezza della produttività del lavoro e della produttività totale dei fattori, unitamente ad altri fattori strutturali, abbia limitato la capacità del sistema produttivo italiano di sostenere la crescita economica e la dinamica delle retribuzioni (Andrews *et al.* 2016; Bugamelli *et al.* 2018). In questo contesto, il tema dei rendimenti salariali dell'istruzione assume particolare rilevanza, considerando il ruolo sempre più centrale del capitale umano e delle competenze nell'accompagnare le transizioni tecnologica, demografica e ambientale che stanno interessando il sistema economico.

I rendimenti dell'istruzione costituiscono da tempo uno dei principali ambiti di studio dell'economia del lavoro. Secondo la teoria del capitale umano (Becker 1964), il premio salariale associato a un maggiore livello di istruzione riflette l'incremento della produttività individuale derivante dall'investimento in formazione. La teoria del *signaling* (Spence 1973), invece, interpreta il titolo di studio come un segnale credibile delle abilità individuali osservabile dai datori di lavoro. Pur muovendo da presupposti differenti, entrambe le impostazioni attribuiscono un ruolo sostanzialmente marginale all'impresa nella determinazione dei rendimenti dell'istruzione: a parità di caratteristiche individuali, lavoratori con lo stesso livello di istruzione dovrebbero ricevere salari simili, indipendentemente dall'impresa presso cui sono occupati.

Negli ultimi due decenni questa interpretazione è stata progressivamente rivista grazie alla crescente disponibilità di basi dati amministrative longitudinali lavoratore-impresa (*linked employer-employee data*), che hanno consentito di osservare simultaneamente le caratteristiche dei lavoratori, delle imprese e delle loro relazioni nel tempo. Abowd *et al.* (1999) hanno introdotto il modello a effetti fissi di lavoratore e impresa (AKM), mostrando come una quota rilevante della variabilità salariale possa essere attribuita a componenti specifiche dell'impresa, oltre che alle caratteristiche individuali. Successivi contributi hanno evidenziato che l'eterogeneità tra imprese rappresenta una determinante fondamentale della dispersione salariale complessiva (Card *et al.* 2013; Song *et al.* 2019). Inoltre, Card *et al.* (2018) mostrano come i processi di *assortative matching* tra lavoratori e imprese contribuiscano significativamente alla disuguaglianza salariale, poiché i lavoratori con caratteristiche produttive migliori tendono a concentrarsi nelle imprese che riconoscono premi salariali più elevati. Questa letteratura suggerisce che anche i rendimenti salariali dell'istruzione possano dipendere, almeno in parte, dalle caratteristiche delle imprese presso cui i lavoratori trovano occupazione. In particolare, l'accesso dei laureati a imprese caratterizzate da premi salariali sistematicamente più elevati potrebbe spiegare una quota dei differenziali retributivi normalmente attribuiti esclusivamente all'investimento in istruzione. In questa prospettiva, i rendimenti dell'istruzione rifletterebbero non soltanto differenze nella produttività individuale, ma anche il diverso accesso a imprese con differenti politiche di wage-setting. L'evidenza empirica più recente si muove in questa direzione. Song *et al.* (2019) mostrano che l'aumento della disuguaglianza salariale negli Stati Uniti è riconducibile in larga misura alla crescente dispersione salariale tra imprese e ai cambiamenti nei meccanismi di allocazione dei lavoratori. Più recentemente, Pan (2025) evidenziano come una quota non trascurabile dei rendimenti dell'istruzione terziaria derivi dall'accesso a imprese che riconoscono premi salariali più elevati, suggerendo che parte dei ritorni dell'investimento educativo rifletta meccanismi di *firm upgrading* piuttosto che esclusivamente differenze di produttività individuale. Nonostante questi progressi, rimane ancora limitata l'evidenza empirica sui meccanismi attraverso cui specifiche caratteristiche delle imprese influenzano i rendimenti salariali dell'istruzione. La letteratura ha infatti documentato il ruolo dei premi salariali d'impresa, ma ha dedicato minore attenzione all'identificazione delle caratteristiche organizzative, istituzionali e produttive che ne sono alla base. Comprendere tali meccanismi appare particolarmente rilevante nel caso italiano, caratterizzato da un'elevata eterogeneità del tessuto produttivo e da persistenti differenze salariali tra imprese.

L'analisi di questi aspetti richiede basi dati in grado di integrare informazioni sui lavoratori, sulle imprese e sulle loro interazioni nel tempo. La crescente disponibilità di dati amministrativi ha ampliato significativamente le possibilità di analisi dei fenomeni economici e sociali. Ciò non implica, tuttavia, che tali archivi siano di per sé sufficienti a rispondere a quesiti di ricerca complessi. Le fonti amministrative sono infatti prodotte prevalentemente per finalità gestionali e regolative e, se considerate isolatamente, presentano limiti informativi che ne restringono il potenziale analitico. Analogamente, le indagini statistiche offrono informazioni dettagliate sulle strategie, sull'organizzazione e sulle relazioni industriali delle imprese, ma non consentono generalmente di osservare con la medesima accuratezza le retribuzioni individuali e le traiettorie lavorative. L'integrazione micro tra dati amministrativi e dati di indagine rappresenta pertanto una strategia fondamentale per superare i limiti delle singole fonti informative e costruire basi dati longitudinali idonee allo studio congiunto delle caratteristiche dei lavoratori, delle imprese e delle loro relazioni.

Muovendosi in questa prospettiva, il presente lavoro analizza in quale misura le caratteristiche delle imprese contribuiscano all'eterogeneità dei rendimenti salariali dell'istruzione universitaria. In particolare, ci si domanda se la diversa capacità delle imprese di riconoscere premi salariali sia influenzata da caratteristiche quali la governance societaria, le relazioni industriali e la redditività aziendale e se tali fattori incidano sui rendimenti salariali dei lavoratori laureati attraverso meccanismi di *rent sharing* o politiche retributive maggiormente favorevoli ai lavoratori ad alta qualificazione.

A tal fine si utilizza una banca dati *employer-*

employee che integra le informazioni amministrative sulle storie lavorative e retributive contenute negli archivi Uniemens (Inps) e nel Sistema delle Comunicazioni Obbligatorie (COB-Mlps) con i dati campionari sulle caratteristiche delle imprese provenienti dalle diverse edizioni dell'Indagine RIL (Inapp). L'applicazione di una strategia econometrica che combina modelli à la Abowd-Kramarz-Margolis (AKM) e regressioni lineari con effetti fissi sui dati Inps-COB-RIL per il periodo 2013-2022 consente di ottenere tre risultati principali. In primo luogo, le politiche retributive delle imprese influenzano in modo positivo seppur debole l'evoluzione dei premi salariali dei laureati essendo questi ultimi riconducibili prevalentemente alla componente legata alle eterogeneità e abilità degli occupati. In secondo luogo, i differenziali retributivi legati al titolo di studio riflettono un processo di selezione (*sorting*) che non è uniforme lungo la distribuzione del capitale umano: l'*assortative matching* tra qualità delle imprese e abilità dei lavoratori è infatti debolmente positivo per i laureati e negativo per i meno qualificati. Infine, le regressioni mostrano come le aziende non trasferiscono automaticamente la performance competitiva in maggiori redditi per i lavoratori più qualificati e quando ciò accade la redistribuzione dei guadagni di produttività avviene solo parzialmente e nel medio periodo.

L'integrazione di tali fonti consente di associare alle caratteristiche organizzative e produttive delle imprese informazioni dettagliate sulle retribuzioni, sulle carriere lavorative e sulla mobilità dei lavoratori, rendendo possibile un'analisi che nessuna delle singole basi informative consentirebbe di realizzare autonomamente.

Il contributo del lavoro è triplice. In primo luogo, fornisce nuova evidenza sul ruolo delle caratteristiche d'impresa nella determinazione dei rendimenti salariali dell'istruzione universitaria, andando oltre le spiegazioni fondate esclusivamente sul capitale umano e sui processi di *sorting*. In secondo luogo, valorizza il potenziale conoscitivo derivante dall'integrazione tra dati amministrativi e dati di indagine, mostrando come tale approccio permetta di collegare gli effetti fissi d'impresa stimati mediante il modello AKM a caratteristiche osservabili delle imprese. Infine, contribuisce al dibattito sulle determinanti microeconomiche della disuguaglianza salariale, evidenziando come le politiche retributive e le caratteristiche organizzative delle imprese influenzino l'efficacia economica dell'investimento in istruzione.

L'articolo è organizzato come segue. La sezione 1 presenta il quadro teorico e la letteratura di riferimento. La sezione 2 descrive il quadro istituzionale che disciplina l'accesso ai dati, le fonti informative utilizzate e il processo di integrazione della banca dati. La sezione 3 illustra la metodologia empirica e discute i principali risultati. La sezione finale conclude.

1. Il quadro teorico e la letteratura di riferimento

Come discusso nell'introduzione, la crescente attenzione al ruolo delle imprese nella determinazione dei salari ha favorito lo sviluppo di due filoni di ricerca strettamente connessi. Il primo analizza il contributo dell'eterogeneità tra imprese alla formazione dei differenziali salariali e della disuguaglianza retributiva; il secondo esamina i meccanismi attraverso cui le caratteristiche economiche e organizzative delle imprese, in particolare la loro capacità di generare rendite, si riflettono sulle retribuzioni attraverso fenomeni di *rent sharing*. È all'intersezione di questi due filoni che si colloca il presente contributo.

Una svolta fondamentale nello studio del ruolo delle imprese nella determinazione dei salari è rappresentata dal modello con effetti fissi additivi di lavoratore e impresa proposto da Abowd *et al.* (1999), che consente di separare empiricamente l'eterogeneità permanente dei lavoratori da quella delle imprese utilizzando dati longitudinali lavoratore-impresa. In questa tradizione, Card *et al.* (2013) mostrano, per la Germania, che l'eterogeneità tra stabilimenti e la sua evoluzione nel tempo spiegano una parte rilevante dell'aumento della disuguaglianza salariale. Grazie all'utilizzo dei dati amministrativi, Song *et al.* (2019) documentano come negli Stati Uniti la crescita della varianza dei guadagni nel lungo periodo sia trainata in larga misura dall'aumento della dispersione tra imprese (oltre che da dinamiche interne alle imprese), consolidando l'idea che la struttura 'firm-level' del mercato del lavoro sia centrale per comprendere la disuguaglianza. Un risultato complementare, nello stesso spirito di decomposizione tra-entro impresa[1] emerge anche da Barth *et al.* (2016), che attribuiscono una parte significativa della crescita della dispersione dei guadagni all'aumento della dispersione tra stabilimenti. Accanto a questi contributi, Bonhomme *et al.* (2019) propongono un quadro *distributional* per dati *matched* che consente una rappresentazione più flessibile dell'eterogeneità tra imprese e del *sorting*.

Rispetto al legame tra eterogeneità tra imprese e rendimenti dell'istruzione, l'idea di fondo è che una

1 Per decomposizione tra-entro le imprese si intende la distinzione tra la componente della dispersione attribuibile alle differenze tra imprese e quella riconducibile alla variabilità entro ciascuna impresa.

parte del premio salariale associato all'istruzione possa riflettere non solo maggiore produttività individuale, ma anche un migliore accesso a imprese che offrono premi salariali più elevati. In questo senso, Pan (2025) fornisce evidenza diretta su questo meccanismo mostrando che una quota dei rendimenti dell'istruzione terziaria può essere attribuita all'accesso a imprese con premi salariali più alti. Questa prospettiva è coerente con la lettura più ampia del ruolo delle imprese nella disuguaglianza e nei salari, sintetizzata da Card *et al.* (2018), che discutono i due canali principali attraverso cui le imprese incidono sui salari tra cui il *sorting* dei lavoratori. Nel nostro contesto, ciò suggerisce che l'eterogeneità dei rendimenti della laurea può emergere sia perché i laureati accedono più spesso a imprese '*better paying*', sia perché a parità di laurea e di lavoratore le imprese differiscono nel modo in cui remunerano le competenze.

Il secondo filone, molto rilevante per la nostra ipotesi di ricerca, studia se e quanto i salari incorporino una componente di condivisione delle rendite generate dall'impresa. Un contributo classico è Van Reenen (1996), che trova evidenza coerente con meccanismi di *rent sharing*, collegando innovazione (e rendite associate) a differenze salariali tra imprese. Kline *et al.* (2019) stimano come shock indotti da brevetti (e quindi da produttività/rendite) si trasmettano ai salari, offrendo evidenza su chi cattura i surplus generati dall'innovazione. In modo complementare, Lamadon *et al.* (2022) sviluppano un modello strutturale con dati *matched* per quantificare l'importanza di rendite e imperfezioni concorrenziali nel mercato del lavoro e discutere le implicazioni per la distribuzione dei salari. Nel loro insieme, questi lavori supportano l'idea che caratteristiche economiche e performance dell'impresa (tra cui la redditività) possano riflettersi nei salari attraverso canali di wage-setting e condivisione delle rendite, e che tali canali possano interagire con la composizione della forza lavoro e con il modo in cui le imprese valorizzano le competenze.

Il presente lavoro si inserisce in questa letteratura combinando due elementi. Da un lato, adottiamo l'impostazione *employer-employee* per stimare e scomporre i premi salariali d'impresa nel solco di Abowd *et al.* (1999) e della successiva letteratura sulla disuguaglianza tra imprese. Dall'altro, riprendiamo l'intuizione di Pan (2025) sul ruolo dei premi delle imprese nei rendimenti di istruzione, concentrandoci su alcune caratteristiche specifiche dell'impresa, fra cui la redditività, come potenziali determinanti osservabili dei premi d'impresa e, quindi, dell'eterogeneità dei rendimenti della laurea.

2. Il quadro istituzionale e le fonti per l'integrazione micro

L'accesso ai dati elementari e il ruolo dell'Inapp nel Sistema statistico nazionale

La possibilità di operare su dati elementari costituisce una condizione imprescindibile per la produzione di analisi quantitative avanzate e per la costruzione di basi informative integrate. Nel caso dell'Inapp, tale possibilità è resa operativa dall'appartenenza dell'Istituto al Sistema statistico nazionale (SISTAN), istituito con il decreto legislativo n. 322 del 6 settembre 1989[2]. Il SISTAN rappresenta la rete dei soggetti pubblici e privati che concorrono alla produzione e alla diffusione dell'informazione statistica ufficiale destinata al Paese e agli organismi internazionali. Ne fanno parte, tra gli altri, l'Istituto nazionale di statistica (Istat), gli uffici di statistica delle amministrazioni centrali e territoriali, le Regioni e Province autonome, le Camere di commercio e altri enti che svolgono funzioni di interesse pubblico in ambito statistico. All'interno del Sistema, l'Istat esercita funzioni di indirizzo, coordinamento e assistenza tecnica, avvalendosi del Comitato di indirizzo e coordinamento dell'informazione statistica (COMSTAT). L'appartenenza al SISTAN non assume una valenza meramente formale, ma comporta prerogative operative rilevanti, tra cui la possibilità di accedere – mediante procedure formalizzate – ad archivi contenenti dati elementari prodotti da altri soggetti del Sistema. Le modalità di scambio dei microdati sono disciplinate dal D.Lgs. n. 322/1989 e da successivi atti di indirizzo, tra cui la Direttiva COMSTAT n. 9 del 2004 e gli aggiornamenti intervenuti nel 2019[3], che hanno precisato condizioni di accesso, finalità ammissibili e garanzie a tutela della riservatezza. Questo assetto normativo e istituzionale consente all'Inapp di operare su microdati, anziché esclusivamente su dati aggregati, nel quadro delle disposizioni del Programma statistico nazionale e nel rispetto della normativa in materia di protezione dei dati personali. La disponibilità di dati elementari rappresenta la condizione abilitante per la costruzione di infrastrutture informative integrate e per la realizzazione di analisi multilivello e longitudinali. Nel caso oggetto del presente contributo, tale prerogativa ha reso possibile

2 Decreto legislativo n. 322 del 6 settembre 1989, *Norme sul Sistema statistico nazionale e sulla riorganizzazione dell'Istituto nazionale di statistica*.

3 Direttiva COMSTAT n. 9/2004; decreto-legge n. 4 del 28 gennaio 2019; decisioni COMSTAT del 2 aprile 2019.

l'integrazione di informazioni relative alle imprese con dati amministrativi riferiti ai lavoratori e ai rapporti di lavoro, permettendo la costruzione di una struttura informativa di tipo *employer-employee*, nella quale le politiche retributive adottate dalle imprese possono essere collegate ai rendimenti dell'istruzione universitaria dei lavoratori.

Le fonti statistiche e amministrative utilizzate

La base dati integrata impiegata nell'analisi empirica deriva dalla combinazione di una rilevazione campionaria condotta dall'Inapp con archivi amministrativi, convertiti a uso statistico, acquisiti tramite il canale SISTAN.

Un ruolo centrale nella costruzione della base dati integrata è svolto dagli archivi amministrativi forniti dall'Inps attraverso il Coordinamento generale statistico attuariale, anch'essi acquisiti mediante canale SISTAN. La fornitura avviene con cadenza annuale e riguarda un campione di individui selezionato secondo il criterio della data di nascita (48 date di nascita per anno), modalità che garantisce la rappresentatività dell'occupazione dipendente nel settore privato non agricolo. La base informativa messa a disposizione comprende differenti archivi amministrativi, riferiti sia a persone sia a imprese: l'anagrafica individuale; le informazioni relative al lavoro dipendente, autonomo e parasubordinato; le misure di sostegno al reddito (indennità di mobilità, disoccupazione, CIG); il casellario dei pensionati; gli estratti conto contributivi; nonché dati riepilogativi relativi alle imprese. Nel contesto del presente contributo, la componente di maggiore rilievo è costituita dai dati relativi ai rapporti di lavoro dipendente afferenti al Fondo lavoratori dipendenti (FLD) nel settore privato extra-agricolo, con esclusione del lavoro domestico. Tali informazioni derivano dalle denunce mensili UNIEmens trasmesse dai datori di lavoro e successivamente rielaborate dall'Inps in modo da produrre record annuali distinti per impresa e caratteristiche contributive del lavoratore.

L'archivio consente di osservare lo stock di occupati con un elevato livello di dettaglio sulle retribuzioni effettivamente corrisposte, sulle settimane lavorate, sulla contribuzione versata e sulle caratteristiche contrattuali. I singoli dati annuali possono essere aggregati in modo da ricostruire un 'rapporto di lavoro' comprensivo dell'intero periodo intercorrente tra data di assunzione e data di cessazione[4], rendendo possibile un'analisi longitudinale coerente con la struttura dei dati SISCO-COB. La rilevanza degli archivi Inps nel presente lavoro è duplice. In primo luogo, essi consentono di disporre di una misura amministrativa della retribuzione, osservata a livello individuale e non soggetta a errori di autovalutazione o dichiarazione. La variabile salariale utilizzata nell'analisi empirica della seconda parte del contributo deriva infatti da queste informazioni contributive. In secondo luogo, la struttura longitudinale degli archivi permette di ricostruire le traiettorie occupazionali dei lavoratori nel tempo, includendo eventuali transizioni tra imprese. Ciò consente di controllare per dinamiche individuali pregresse e per effetti di mobilità, elementi fondamentali per stimare correttamente la dinamica salariale individuale e per isolare una componente residua della retribuzione, successivamente utilizzabile in analisi a livello d'impresa.

La seconda banca dati amministrativa è rappresentata dal Sistema delle Comunicazioni Obbligatorie (COB) online, un archivio gestito dal Ministero del Lavoro e delle politiche sociali, che raccoglie informazioni sugli eventi che caratterizzano i rapporti di lavoro: instaurazioni, trasformazioni, proroghe e cessazioni. L'obbligo di comunicazione telematica è stato introdotto dalla legge n. 296/2006[5]. L'archivio ha natura di flusso: l'unità elementare è l'evento amministrativo, che può essere aggregato per ricostruire il 'rapporto di lavoro' intercorrente tra la comunicazione di avviamento e quella di cessazione. L'archivio copre la quasi totalità del lavoro dipendente e alcune tipologie di lavoro autonomo e parasubordinato. Occorre, tuttavia, considerare alcuni limiti strutturali: i lavoratori assunti prima dell'introduzione dell'obbligo di comunicazione e che non hanno subito variazioni nel rapporto di lavoro possono non comparire nell'archivio[6]; inoltre, i dati sono disponibili con un certo ritardo temporale rispetto al periodo di riferimento[7]. Ai fini delle analisi presentate nel presente articolo, SISCO svolge una funzione complementare rispetto agli archivi previdenziali, fornendo informazioni sulle qualifiche professionali e sui titoli di studio dei lavoratori, non presenti negli archivi Inps.

Con riferimento alla banca dati di natura campionaria, prima di descrivere nel dettaglio RIL è necessario introdurre il Registro statistico

4 I singoli record annuali possono essere aggregati in modo da ricostruire un 'rapporto di lavoro' che, analogamente a quanto avviene per SISCO, comprende tutte le informazioni intercorrenti tra la data di assunzione e quella dell'eventuale cessazione.

5 Legge n. 296/2006 (Legge Finanziaria 2007).

6 Lavoratori assunti prima del 2008 e senza successive variazioni contrattuali possono non risultare nell'archivio.

7 I dati sono disponibili con uno scarto temporale di circa due trimestri rispetto al periodo di riferimento.

delle imprese (ASIA) dell'Istat che ne costituisce la base informativa per la definizione dell'universo di riferimento e per il disegno campionario, probabilistico e complesso garantendo la coerenza tra la popolazione osservata nella survey e quella rilevata nei registri statistici ufficiali. ASIA include le unità economiche che esercitano arti e professioni nelle attività industriali, commerciali e dei servizi, fornendo informazioni anagrafiche e strutturali quali attività economica, numero di addetti, forma giuridica, data di inizio e fine attività e fatturato. Dal campo di osservazione sono escluse le attività relative all'agricoltura, alla pubblica amministrazione, alle organizzazioni associative e ad altri settori specifici[8]. Tale perimetro risulta coerente con quello adottato nella RIL e con la copertura dei principali archivi amministrativi sul lavoro dipendente nel settore privato. La funzione di ASIA nel presente contesto non è soltanto quella di frame campionario, ma anche quella di riferimento strutturale per garantire coerenza tra l'universo delle imprese osservate nella survey e quelle presenti negli archivi amministrativi.

La Rilevazione longitudinale su Imprese e Lavoro-RIL costituisce la fonte campionaria utilizzata nell'analisi. Si tratta di un'indagine statistica inserita nel Programma statistico nazionale (PSN) e caratterizzata da obbligo di risposta per i soggetti privati coinvolti, in quanto rilevazione di interesse pubblico. La prima edizione della RIL risale al 2005; l'indagine è stata successivamente riproposta con cadenza periodica, fino all'edizione più recente (la VII) attualmente in fase di realizzazione. La continuità nel tempo della rilevazione consente di disporre di una base informativa strutturata e coerente, idonea ad analisi diacroniche sui comportamenti delle imprese. L'indagine è finalizzata allo studio del comportamento delle imprese nei confronti del lavoro, con particolare riferimento alle strategie organizzative, alle politiche di gestione delle risorse umane, alle relazioni industriali e all'introduzione di innovazioni tecnologiche e organizzative. Le informazioni raccolte consentono di analizzare le relazioni tra le scelte aziendali in materia di utilizzo del lavoro e i risultati economici dell'impresa. L'unità di analisi è l'impresa. Il campione, composto da circa 30.000 unità, è rappresentativo delle società di capitali e delle società di persone attive sul territorio nazionale nei settori privati non agricoli. Sono pertanto escluse dal campo di osservazione le imprese individuali e le altre forme giuridiche non rientranti in tali categorie, oltre ai settori economici non compresi nel perimetro definito dal Registro ASIA[9]. Un elemento metodologicamente rilevante della RIL è la presenza di una componente longitudinale: una quota, circa il 30%, del campione è costituita da imprese già intervistate nelle edizioni precedenti, consentendo di seguire nel tempo l'evoluzione delle strategie aziendali. Nel contesto dell'analisi applicativa sviluppata nella seconda parte del contributo, la RIL fornisce l'insieme di variabili relative a redditività, relazioni industriali, investimenti, innovazione e competitività. L'integrazione con gli archivi Inps e SISCO consente di collegare tali informazioni alle retribuzioni effettivamente percepite dai lavoratori impiegati nelle imprese osservate, costruendo una base dati multilivello nella quale il livello decisionale dell'impresa può essere messo in relazione con gli esiti salariali individuali nel tempo costruendo una struttura *employer-employee*. Le edizioni utilizzate nell'articolo fanno riferimento alle rilevazioni condotte nel 2015, 2018 e 2022.

Costruzione della struttura employer-employee panel

L'obiettivo dell'integrazione tra RIL, archivi Inps e SISCO è la costruzione di una struttura informativa di tipo *employer-employee*, nella quale le caratteristiche e le scelte dell'impresa possano essere collegate agli esiti individuali dei lavoratori nel tempo. A tal fine, il primo passaggio consiste nell'identificazione delle imprese incluse nella RIL e nel collegamento di tali unità con gli archivi amministrativi mediante il codice fiscale. Una volta individuate le imprese della survey negli archivi Inps e SISCO, vengono selezionati tutti i lavoratori che risultano aver intrattenuto un rapporto di lavoro con tali imprese nel periodo di osservazione. In un secondo momento, per ciascun lavoratore così identificato, viene ricostruita l'intera storia lavorativa disponibile negli archivi amministrativi, indipendentemente dal datore di lavoro. La struttura risultante è dunque un panel multilivello nel quale l'impresa costituisce il livello decisionale e il lavoratore rappresenta l'unità di outcome.

La costruzione della base dati integrata richiede un'articolata attività di verifica e validazione, finalizzata a garantire la coerenza interna degli

8 Sono escluse, tra le altre, le attività relative ad agricoltura (sezione A), amministrazione pubblica (sezione O), organizzazioni associative (divisione 94), famiglie come datori di lavoro (sezione T), organismi extraterritoriali (sezione U).

9 Sono escluse dal campo di osservazione le attività economiche relative a: agricoltura, silvicoltura e pesca (sezione A); amministrazione pubblica e difesa; assicurazione sociale obbligatoria (sezione O); attività di organizzazioni associative (divisione 94); attività di famiglie e convivenze come datori di lavoro per personale domestico; produzione di beni e servizi indifferenziati per uso proprio da parte di famiglie e convivenze (sezione T); organizzazioni ed organismi extraterritoriali (sezione U); unità classificate come istituzioni pubbliche e istituzioni private non profit.

archivi e l'affidabilità delle informazioni risultanti dall'integrazione. Le verifiche riguardano, in primo luogo, la coerenza temporale dei rapporti di lavoro e delle retribuzioni osservate, con particolare attenzione alla continuità delle carriere individuali e alla corretta attribuzione dei lavoratori alle imprese nel tempo. In secondo luogo, vengono effettuati controlli incrociati sulle variabili comuni tra le diverse fonti, al fine di identificare eventuali discrepanze e valutarne l'impatto sull'analisi. È inoltre necessario tenere conto dei limiti derivanti dalla parziale sovrapposizione degli universi osservati, dalla natura campionaria della fornitura Inps e dalle possibili lacune informative presenti negli archivi amministrativi. Tali elementi vengono considerati nella definizione dell'universo analitico e nell'interpretazione dei risultati. Nonostante tali limiti, l'integrazione tra le fonti consente di ampliare significativamente il contenuto informativo disponibile rispetto all'utilizzo isolato delle singole basi dati, permettendo di costruire una struttura *employer-employee* panel coerente e metodologicamente solida[10].

3. Analisi empirica

La strategia econometrica

In questo paragrafo illustriamo la strategia econometrica adottata per analizzare in quale misura le caratteristiche delle imprese contribuiscano all'eterogeneità dei rendimenti salariali dell'istruzione universitaria. L'analisi si articola in due fasi sequenziali e complementari. Nella prima vengono stimati i premi salariali d'impresa attraverso un modello a effetti fissi lavoratore-impresa (AKM), distinguendo tra lavoratori laureati e non laureati; nella seconda tali premi salariali sono messi in relazione alle caratteristiche economiche, organizzative e istituzionali delle imprese rilevate dall'indagine RIL. La prima fase dell'analisi si colloca nel solco della letteratura inaugurata da Abowd *et al.* (1999), secondo cui la determinazione dei salari riflette congiuntamente l'eterogeneità osservata e non osservata dei lavoratori e quella delle imprese presso cui essi sono occupati (Card *et al.* 2013; Bonhomme *et al.* 2019 ; Casarico e Lattanzio 2024). Utilizziamo i dati amministrativi integrati Inps-COB per stimare una versione estesa del modello AKM nella quale gli effetti fissi individuali sono invarianti nel tempo, mentre la componente di impresa è consentita variare tra differenti intervalli temporali, così da ricostruire l'evoluzione delle politiche retributive aziendali (Babet *et al.* 2015; Lachowska *et al.* 2023).

Coerentemente con le indicazioni della letteratura più recente, il modello viene stimato separatamente per due sottocampioni di lavoratori: quelli in possesso di un titolo di istruzione terziaria e quelli con un titolo di studio non universitario. La classificazione è definita in modo costante nel tempo. In particolare, il livello di istruzione è identificato mediante una variabile dummy che assume valore pari a uno per i lavoratori laureati e zero altrimenti, costruita sulla base del massimo livello di istruzione osservato per ciascun individuo nell'intero periodo disponibile. Tale scelta evita che eventuali variazioni nella registrazione del titolo di studio interferiscano con l'identificazione degli effetti fissi individuali e consente un confronto omogeneo tra lavoratori laureati e non laureati lungo tutto l'orizzonte temporale di osservazione. Formalmente, per ciascun gruppo di istruzione (e) (laureati e non laureati) viene stimata la seguente equazione:

$$[1] \qquad \ln w_{ijt,e} = \alpha_{i,e} + \psi_{jt,e} + X_{it,e}\beta + \varepsilon_{it,e}$$

dove $\ln w_{ijt,e}$ rappresenta il logaritmo del salario settimanale full-time equivalent del lavoratore (i), occupato nell'impresa (j) nel periodo t = [2013-16 2017-19 2020-22]. Il vettore $X_{it,e}$ include le principali caratteristiche osservabili del lavoratore, mentre rappresenta l'effetto fisso individuale e cattura l'eterogeneità non osservata e invariante nel tempo. Il termine $\psi_{jt,e}$ identifica l'effetto fisso impresa-periodo e misura il premio salariale riconosciuto dall'impresa (j) ai lavoratori appartenenti al gruppo di istruzione (e) nel periodo (t). Il termine di errore $\varepsilon_{it,e}$ è un disturbo idiosincratico a media nulla e varianza finita. La differenza tra i premi salariali stimati per laureati e non laureati, $(\psi^{L}_{jt} - \psi^{NL}_{jt})$, rappresenta la componente *firm-specific* dei rendimenti salariali dell'istruzione universitaria e costituisce la variabile dipendente dell'analisi svolta nel secondo stadio.

Nella seconda fase le stime ottenute dal modello AKM sono integrate con le informazioni raccolte da RIL, al fine di verificare se e in quale misura la dinamica dei premi salariali riconosciuti ai lavoratori laureati sia associata alle caratteristiche economiche, organizzative e istituzionali delle imprese. A tale scopo viene stimato il seguente modello lineare:

$$[2] \quad \psi^{L\text{-}NL}_{jt} = \beta_0 + \beta_1 \ln(\Pi_{jt}) + \beta_2 \ln(\Pi_{jt}) \times 1(2017\text{–}19) + \beta_3 \ln(\Pi_{jt}) \times 1(2020\text{–}22) + \delta' F_{jt} + \mu_j + \varepsilon_{jt}$$

10 Si rimanda all'Appendice metodologica per una descrizione approfondita delle procedure di pretrattamento e standardizzazione delle forniture, *record linkage* e integrazione micro tra archivi.

dove $\psi_{jt}^{L\text{-}NL}$ rappresenta il differenziale tra i premi salariali stimati per laureati e non laureati nell'impresa (j) e nel periodo (t). Tra le variabili di controllo di fonte RIL focalizziamo l'attenzione sul (log del) fatturato annuo per dipendente Π_{jt} e sui termini di interazione della redditività con le variabili indicatrici degli intervalli di tempo che – a loro volta – possono essere interpretati come una sorta di elasticità 'dinamica tra performance e politiche salariali per l'istruzione'[11]. Il vettore F_{jt} include le caratteristiche del management, della specializzazione produttiva, l'organizzazione del personale, relazioni industriali, altro. Il parametro μ_j identifica l'eterogeneità non osservata e invariante nel tempo, mentre ε_{jt} è un disturbo idiosincratico. È opportuno sottolineare come la strategia econometrica viene declinata nella struttura statistico informativa del nostro dataset – anche per offrire un ordito interpretativa dei risultati delle regressioni. I premi salariali d'impresa vengono infatti stimati separatamente nei tre sottoperiodi 2013-2016, 2017-2019 e 2020-2022, costruiti in corrispondenza delle tre edizioni della RIL (2015, 2018 e 2021). Tale scelta consente di associare a ciascun premio salariale le caratteristiche dell'impresa osservate nella corrispondente rilevazione RIL, garantendo la coerenza temporale tra le variabili dipendenti stimate nel primo stadio e le covariate utilizzate nel secondo.

Le stime del modello AKM sono effettuate sulla massima componente connessa lavoratore-impresa (*largest connected set*[12]), così da garantire l'identificazione degli effetti fissi d'impresa e la confrontabilità dei premi salariali stimati all'interno di ciascun sottocampione di istruzione. L'utilizzo della massima componente connessa contribuisce inoltre a ridurre i problemi di identificazione derivanti dalla limitata mobilità dei lavoratori tra imprese (*limited mobility bias*), pur senza eliminarli completamente; di conseguenza, i risultati della decomposizione della varianza devono essere interpretati soprattutto in termini comparativi tra lavoratori laureati e non laureati.

Le elaborazioni econometriche fanno, infine, riferimento a un campione finale di lavoratori dipendenti di età compresa tra 20 e 60 anni occupati in imprese private extra-agricole con almeno dieci dipendenti e per i quali è disponibile almeno un rapporto di lavoro nelle COB, fonte dalla quale viene ricavata l'informazione sul titolo di studio. Tale selezione consente di concentrare l'analisi su imprese caratterizzate da una minima organizzazione dei mercati interni del lavoro e, quindi, potenzialmente in grado di adottare politiche retributive differenziate tra gruppi di lavoratori. L'analisi empirica è condotta su un campione di oltre 10 milioni di osservazioni, corrispondenti a circa 2 milioni di lavoratori, dei quali circa 400 mila in possesso di un titolo universitario.

I risultati principali

La tabella 1 riporta le principali statistiche descrittive del campione, distinguendo tra lavoratori laureati e non laureati, sia con riferimento all'intero campione sia alla restrizione campionaria corrispondente al *largest connected set* lavoratore-impresa (utilizzata per la stima del modello AKM).

Le statistiche descrittive evidenziano differenze sistematiche tra i due gruppi di lavoratori. Nel campione completo, i laureati percepiscono una retribuzione settimanale mediamente più elevata rispetto ai non laureati (pari rispettivamente a 6,4 e 6,2), risultano mediamente più giovani (37 contro 41 anni circa) e presentano una maggiore incidenza della componente femminile (56% contro 39%). Differenze pressoché identiche si osservano nel *largest connected set*, indicando che la restrizione del campione necessaria all'identificazione degli effetti fissi lavoratore-impresa non altera in misura significativa le caratteristiche osservabili dei lavoratori né il differenziale salariale tra i due gruppi. Nel complesso, tali evidenze risultano coerenti con la letteratura sul mercato del lavoro italiano, che documenta l'esistenza di un premio salariale associato al conseguimento di un titolo universitario, sebbene di entità relativamente contenuta e in progressiva riduzione rispetto a quanto osservato nella maggior parte dei Paesi OCSE (Naticchioni *et al.* 2010).

Il semplice confronto delle retribuzioni medie, tuttavia, non consente di stabilire in quale misura tale differenziale rifletta la maggior dotazione di capitale umano di chi è in possesso di un titolo universitario oppure sia riconducibile alle specifiche caratteristiche e comportamenti delle aziende presso cui sono occupati. In particolare, una quota del differenziale retributivo associato alla laurea potrebbe essere correlata a meccanismi di *sorting* e processo di *assortative matching*: gli individui con maggiori abilità non osservate tendono ad essere concentrati in tipologie di imprese che adottano politiche più favorevoli alla forza lavoro qualificata. E il processo di *assortative matching* potrebbe essere intrinsecamente differente, nel segno e nel valore

11 Nel dettaglio, il coefficiente β_1 misura il legame tra redditività e premio per laureati nel periodo iniziale (2013-2016), mentre i coefficienti β_2 e β_3 indicano come evolve tale nesso nel 2017-2019 e nel 2020-2022, rispettivamente.

12 La massima componente connessa è identificata come il più ampio sotto grafo del network lavoratore-impresa nel quale ciascuna impresa risulta collegata alle altre attraverso almeno una transizione lavorativa osservata.

Tabella 1. Statistiche descrittive

	Campione totale		*Largest connected set*	
	Mean	Std. dev.	Mean	Std. dev.
		Laureati		
Log (salario)	6,415	0,541	6,421	0,542
Età (in anni)	37,18	9,148	37,13	9,111
Donne	0,563	0,496	0,567	0,496
N. osservazioni	2.102.115		1.967.173	
		Non laureati		
Log (salario)	6,195	0,426	6,194	0,426
Età (in anni)	40,63	11,22	40,54	11,21
Donne	0,387	0,487	0,386	0,486
N. osservazioni	8.442.235		8.039.660	

Fonte: elaborazioni degli Autori su dati Inps-COB 2013-2022

assoluto, lungo la distribuzione del capitale umano.

L'analisi econometrica sviluppata nei paragrafi successivi è finalizzata proprio a distinguere queste due componenti. Attraverso la stima del modello AKM separatamente per lavoratori laureati e non laureati è infatti possibile isolare il contributo della componente *firm-specific* alla determinazione dei salari e verificare successivamente se le differenze nei premi salariali riconosciuti ai lavoratori con istruzione universitaria risultino sistematicamente associate alle caratteristiche economiche, organizzative e istituzionali delle imprese. In questa prospettiva, la tabella 2 riporta le stime del modello AKM ottenute separatamente per i due gruppi di lavoratori. La specificazione consente di isolare il contributo dell'eterogeneità individuale e di quella d'impresa alla determinazione dei salari all'interno dei due gruppi di istruzione, fornendo la base per analizzare il ruolo delle caratteristiche aziendali nella formazione dei rendimenti salariali dell'istruzione. Per entrambi i sottocampioni, i coefficienti associati all'età e al suo quadrato presentano rispettivamente segno positivo e negativo, delineando un profilo salariale concavo lungo il ciclo di vita lavorativa, coerentemente con le previsioni della teoria del capitale umano. L'effetto risulta più pronunciato tra i laureati, suggerendo una maggiore crescita delle retribuzioni nelle fasi iniziali e centrali della carriera, in linea con percorsi professionali caratterizzati da una più elevata accumulazione di capitale umano specifico e da maggiori opportunità di progressione salariale.

L'inclusione degli effetti fissi individuali (α_i) e degli effetti fissi impresa-periodo (ψ_{jt}) consente di spiegare una quota molto elevata della variabilità salariale osservata (R^2 pari a 0,87 per i laureati e 0,83 per i non laureati), confermando l'elevata capacità esplicativa del modello AKM applicato a dati amministrativi lavoratore–impresa. Le stime degli effetti fissi ottenute in questa fase costituiscono la base per analizzare il contributo relativo delle componenti individuali e aziendali alla dispersione salariale.

A questo punto è possibile esaminare i fattori sottostanti la dispersione salariale e la natura del processo di *matching/sorting* tra individui e caratteristiche delle imprese. La tabella 3 riporta la decomposizione della varianza e la matrice varianza-covarianza delle componenti stimate dal modello AKM. In linea con la letteratura sui dati *matched employer-employee* (Card *et al.* 2013), i risultati evidenziano che la dispersione salariale è spiegata prevalentemente dall'eterogeneità non osservata dei lavoratori, mentre gli effetti fissi di impresa, pur significativi contribuiscono in misura più contenuta alla variabilità complessiva delle retribuzioni. Questo risultato emerge per entrambi i sottocampioni confermando che le caratteristiche permanenti dei lavoratori rappresentano il principale fattore di differenziazione delle retribuzioni, mentre le politiche salariali delle imprese contribuiscono a spiegare una quota più contenuta della dispersione osservata[13].

13 Per i lavoratori laureati la varianza associata agli effetti fissi individuali ($Var(\alpha_i) = 0,146$) risulta sensibilmente superiore a quella attribuibile agli effetti fissi d'impresa ($Var(\psi_{jt}) = 0,078$); una configurazione analoga si osserva anche tra i non laureati, per i quali la componente individuale continua a rappresentare la principale fonte di eterogeneità salariale ($Var(\alpha_i) = 0,099$ contro $Var(\psi_{jt}) = 0,070$).

Tabella 2. Stime AKM. Variabile dipendente (log) salario settimanale

	Laureati	Non laureati
Età	0,083***	0,041***
	[0,003]	[0,001]
Età^2	[0,001***	-0,000***
	[0,000]	[0,000]
$\alpha_{i,e}$	Sì	Sì
$\psi_{jt,e}$	Sì	Sì
Constante	4.377***	5.108***
	[0,036]	[0,021]
N. lavoratori	373.726	1.575.682
N. imprese*periodo	160.933	567.912
N. osservazioni	1.967.173	8.039.660
R2	0,87	0,83

Note: le regressioni non includono variabili di controllo per le caratteristiche demografiche che risultano invarianti nel tempo. Errori standard robusti clusterizzati a livello di impresa sono riportati in parentesi. Significatività statistica: *** al 1%, ** al 5%, al 10%.
Fonte: elaborazione degli Autori su dati campione Inps-COB

Tabella 3. Scomposizione varianza-covarianza stime AKM

	α_i	ψ_{jt}	$X_{it}\,\beta$	Ln(salario)
		Laureati		
α_i	0,146			
ψ_{jt}	0,006	0,078		
$X_{it}\,\beta$	-0,019	0,003	0,062	
Ln(salario)	0,133	0,087	0,047	0,294
		Non laureati		
α_i	0,099			
ψ_{jt}	-0,007	0,070		
$X_{it}\,\beta$	-0,015	0,002	0,028	
Ln(salario)	0,077	0,065	0,016	0,181

Fonte: elaborazioni degli Autori su dati Inps-COB 2013-2022

La matrice varianza-covarianza consente, inoltre, di approfondire il processo di allocazione dei lavoratori tra imprese. Per i lavoratori con istruzione universitaria emerge una covarianza positiva, seppure di entità molto contenuta, tra effetto fisso individuale ed effetto fisso d'impresa ($Cov(\alpha_i, \psi_{jt})$ = 0,006). Tale evidenza suggerisce la presenza di *assortative matching* positivo, ossia una moderata tendenza dei lavoratori caratterizzati da una maggiore componente di abilità non osservata a essere occupati presso imprese che riconoscono premi salariali relativamente più elevati. Il risultato è coerente sia con la letteratura internazionale sui dati *matched employer-employee*, che documenta l'esistenza di processi di allocazione tra lavoratori e imprese in grado di influenzare la dispersione salariale (Card *et al.* 2013), sia con la più recente evidenza riferita al mercato del lavoro italiano, secondo cui una parte dei differenziali retributivi dei laureati riflette il processo di *sorting* verso imprese caratterizzate da differenti politiche salariali (Bonacini *et al.* 2025). La ridotta entità della covarianza osservata suggerisce, tuttavia, che tale meccanismo, pur presente, non rappresenti l'unico fattore alla base dei rendimenti salariali dell'istruzione universitaria. Per i lavoratori non laureati si osserva invece una covarianza lievemente

negativa tra effetto fisso individuale ed effetto fisso d'impresa ($Cov(\alpha_i, \psi_{jt}) = -0{,}007$), che segnala un processo di allocazione di segno opposto rispetto a quello osservato per i laureati. Sebbene anche in questo caso la magnitudine della covarianza sia estremamente contenuta, la diversa configurazione dei due sottocampioni suggerisce che i meccanismi di matching tra lavoratori e imprese operino in modo differenziato a seconda del livello di istruzione[14]. In sintesi, quanto emerge nella tabella 3 conferma l'idea che il ruolo delle politiche retributive a livello aziendale non costituiscono la principale fonte della dispersione salariale complessiva. Ciò non implica che esse siano economicamente irrilevanti. Al contrario, la circostanza che la componente *firm-specific* presenti un peso relativamente contenuto nella varianza aggregata rende particolarmente interessante comprendere quali caratteristiche delle imprese siano associate alla sua eterogeneità. In altre parole, se i differenziali salariali riconosciuti ai lavoratori laureati non dipendono esclusivamente dalla composizione della forza lavoro diventa naturale chiedersi se vi siano fattori microeconomici interni al ciclo di vita aziendale (le relazioni industriali, la redditività e la capacità di generare rendite) che contribuiscano a spiegare le differenze nei premi salariali riconosciuti ai lavoratori più qualificati.

È proprio questa l'ipotesi che viene verificata nel secondo stadio dell'analisi. La tabella 4 riporta i risultati delle stime dell'equazione [2], ottenute sulla componente longitudinale della banca dati integrata Inps-COB-RIL, con l'obiettivo di verificare se le caratteristiche economiche e organizzative delle imprese contribuiscano a spiegare l'eterogeneità dei premi salariali riconosciuti ai lavoratori laureati. Nello specifico, la colonna (1) presenta le stime ottenute mediante regressione OLS. I coefficienti associati alla performance economica dell'impresa, misurata attraverso il logaritmo del fatturato per dipendente, non risultano statisticamente significativi né nel periodo iniziale né nei successivi intervalli temporali. Considerata isolatamente, tale evidenza sembrerebbe suggerire l'assenza di una relazione sistematica tra performance aziendale e premi salariali riconosciuti ai lavoratori con istruzione universitaria. Tale interpretazione deve essere considerata con cautela, poiché le stime OLS non tengono conto dell'eterogeneità non osservata tra imprese, che può riflettere caratteristiche permanenti quali modelli di governance, pratiche manageriali, assetti organizzativi e altre determinanti strutturali delle politiche retributive. La colonna (2), che prevede l'introduzione nel modello degli effetti fissi di impresa, restituisce un quadro sensibilmente diverso: una volta controllato per l'eterogeneità aziendale invariante nel tempo emerge infatti una relazione dinamica tra performance economica e premi salariali riconosciuti ai lavoratori laureati.

Nel periodo iniziale (2013-2016), un incremento del fatturato per dipendente risulta associato a una riduzione del differenziale salariale a favore dei laureati (-0,9%). Tale risultato non implica che i laureati percepiscano retribuzioni inferiori rispetto ai lavoratori non laureati; suggerisce piuttosto che, nelle imprese caratterizzate da una migliore performance economica, il premio salariale associato all'istruzione universitaria risulti relativamente più contenuto. Nei periodi successivi il quadro empirico evolve progressivamente. L'interazione tra performance economica e periodo 2017-2019 assume segno positivo, pur rimanendo statisticamente non significativa, mentre nel periodo 2020-2022 il coefficiente diviene positivo e statisticamente significativo. Nell'ultima fase del periodo analizzato, pertanto, una migliore performance economica dell'impresa risulta associata a un incremento dei premi salariali riconosciuti ai lavoratori laureati, suggerendo che una quota crescente dei benefici economici generati dall'impresa venga destinata alla componente più qualificata della forza lavoro. I risultati indicano, quindi, che la relazione tra performance aziendale e rendimenti salariali dell'istruzione non è stabile nel tempo, ma evolve nel corso del periodo osservato. Il progressivo rafforzamento del legame nell'ultima fase suggerisce che gli effetti della maggiore performance economica sulle politiche retributive possano manifestarsi gradualmente, nell'ambito dei meccanismi di determinazione dei salari propri del sistema italiano di relazioni industriali. Tale interpretazione appare coerente con la letteratura che evidenzia come le istituzioni della contrattazione collettiva contribuiscano a modellare la trasmissione dei differenziali di produttività ai salari e a contenere la dispersione delle politiche retributive tra imprese (Devicienti *et al.* 2019).

14 Nell'interpretare il segno e la magnitudine della covarianza tra effetti fissi individuali ed effetti fissi d'impresa è opportuno considerare le possibili implicazioni del cosiddetto *limited mobility bias*, ampiamente discusso nella letteratura sui modelli AKM (Saggio *et al*., 2023). Tale fenomeno deriva dalla limitata mobilità dei lavoratori tra imprese, dalla quale dipende l'identificazione degli effetti fissi d'impresa, e può influenzare la stima della componente *firm-specific* e, conseguentemente, della covarianza tra lavoratori e imprese. Pertanto, i risultati relativi al processo di assortative matching devono essere interpretati principalmente in termini comparativi tra i due gruppi di lavoratori, piuttosto che come una misura puntuale dell'intensità del matching.

Tabella 4. Stime regressioni lineari. Variabile dipendente: premi salariali laureati (vs non laureati) per impresa e periodo

	[1]	[2]
$\ln(\Pi_{jt}) \times 1(2020–22)$	0,007	0,008*
	[0,006]	[0,005]
$\ln(\Pi_{jt}) \times 1(2017–19)$	0,002	0,003
	[0,006]	[0,004]
$\ln(\Pi_{jt})$	-0,008	-0,009**
	[0,005]	[0,004]
Vacancy ratio	0,029	0,056*
	[0,020]	[0,034]
Associazione datoriale	0,029*	0,012
	[0,015]	[0,010]
Altri controlli	Sì	Sì
Effetti fissi impresa	No	Sì
Costante	0,020	0,038
	[0,054]	[0,040]
N. imprese	1.247	1.247
N. osservazioni	3.502	3.502
R2	0,027	0,752

Note: altri controlli includono livello di istruzione e genere del management, governance familiare, numero di dipendenti (in classi dimensionali), età (in anni) dell'impresa. Tutte le regressioni includono effetti fissi per regione geografica e settore di attività (ATECO 2 digit). Errori standard clusterizzati a livello di impresa riportati in parentesi. Significatività statistica: *** al 1%, ** al 5%, al 10%.
Fonte: elaborazioni degli Autori su dati Inps-COB-RIL

Questi risultati sembrano essere coerenti con l'ipotesi che le caratteristiche delle imprese contribuiscano, almeno in parte, a spiegare l'eterogeneità dei rendimenti salariali dell'istruzione universitaria. Sebbene la componente *firm-specific* rappresenti una quota relativamente contenuta della dispersione salariale complessiva, i premi salariali riconosciuti ai lavoratori laureati risultano sistematicamente associati ad alcune caratteristiche economiche delle imprese nelle quali essi sono occupati quali ad esempio un diverso contesto organizzativo e retributivo. Tale evidenza è in linea con la recente letteratura che attribuisce alle imprese un ruolo non trascurabile nella determinazione della disuguaglianza salariale e dei rendimenti dell'istruzione, evidenziando come le politiche di *wage-setting* e l'eterogeneità tra imprese possano contribuire ai differenziali retributivi osservati (Card *et al.* 2018; Bonacini *et al.* 2025; Pan 2025).

Conclusioni

Le analisi presentate nelle pagine precedenti hanno esaminato la relazione tra politiche salariali delle imprese e rendimento dell'istruzione universitaria, considerando congiuntamente l'eterogeneità degli individui, le caratteristiche delle aziende in cui sono occupati e il processo di matching tra domanda e offerta di lavoro qualificato.

A tal fine è stato sviluppato un sistema informativo integrato che combina le informazioni sulle storie retributive e lavorative contenute negli archivi Uniemens (Inps) e nel Sistema delle Comunicazioni Obbligatorie con i dati campionari sulle imprese raccolti dall'Indagine RIL (Inapp).

Le stime econometriche restituiscono evidenze relativamente nuove per il mercato del lavoro italiano. In primo luogo, emerge un ruolo positivo, seppur contenuto, delle politiche retributive aziendali nel condizionare l'andamento dei premi salariali dei laureati, che restano comunque influenzati soprattutto dalla componente legata alle abilità e all'eterogeneità individuale. Inoltre, i differenziali retributivi associati al titolo di studio riflettono un processo di selezione non uniforme lungo la distribuzione del capitale umano: l'*assortative matching* tra qualità delle imprese e abilità dei lavoratori risulta debolmente positivo per i laureati e negativo per gli occupati meno qualificati.

Le regressioni basate sui dati d'impresa mostrano poi che le aziende italiane non trasformano automaticamente la propria performance competitiva in salari più elevati per i lavoratori altamente qualificati. Quando la redistribuzione dei guadagni di produttività si verifica, essa è parziale e si manifesta solo nel medio periodo. Tale risultato suggerisce che il tessuto imprenditoriale contribuisce alla compressione dei rendimenti salariali dell'istruzione, indipendentemente dall'azione di eventuali meccanismi compensativi legati agli assetti delle relazioni industriali o alla regolamentazione dei contratti collettivi.

Nel complesso, i premi dell'istruzione non possono essere interpretati come un mero riflesso della produttività individuale: dipendono anche dalla collocazione dei lavoratori in imprese che risultano eterogenee per strategie, istituzioni e capacità di generare premi salariali. Ne deriva che le politiche educative e del lavoro dovrebbero considerare non solo l'accumulazione di capitale umano, ma anche le condizioni che favoriscono l'accesso dei laureati alle imprese di qualità e a contesti capaci di valorizzarne le competenze.

In conclusione, il nostro studio contribuisce alla letteratura su capitale umano, disuguaglianza salariale ed eterogeneità d'impresa, mostrando che una piena comprensione dei rendimenti dell'istruzione richiede un'analisi congiunta delle strategie aziendali e delle traiettorie lavorative individuali.

Appendice metodologica

Pretrattamento e standardizzazione delle forniture

L'attività di acquisizione delle banche dati contenenti informazioni elementari è svolta, per Inapp, dal Servizio statistico, che provvede, una volta ricevute le forniture, all'implementazione delle necessarie procedure di pretrattamento e trattamento dei dati. Tali attività sono finalizzate alla predisposizione di dataset controllati, coerenti e pseudonimizzati, idonei all'elaborazione statistica da parte del personale autorizzato. Le procedure di trattamento si differenziano in funzione della natura della fonte, delle modalità di raccolta originaria delle informazioni e del grado di strutturazione delle basi dati fornite.

Nel caso delle forniture derivanti dal Sistema delle Comunicazioni Obbligatorie (SISCO), i dati trasmessi dal Ministero del Lavoro sono già oggetto di una prima rielaborazione e vengono forniti in forma strutturata come 'rapporti di lavoro', ossia come aggregazioni degli eventi amministrativi intercorsi tra l'avvio e l'eventuale cessazione del rapporto. Tuttavia, nonostante tale trattamento preliminare, la base dati mantiene un elevato grado di complessità. La natura amministrativa dell'archivio, la numerosità dei record e la possibile frammentazione o sovrapposizione dei rapporti nel tempo rendono necessario un ulteriore intervento di controllo e armonizzazione. In particolare, è indispensabile verificare la coerenza temporale dei rapporti ricostruiti, garantire la congruenza tra identificativi di impresa e lavoratore, armonizzare le classificazioni professionali e i titoli di studio e assicurare l'allineamento con le variabili presenti negli archivi Inps e nella RIL.

Per quanto riguarda le forniture Inps, il livello di strutturazione è più elevato dato un primo trattamento statistico realizzato a monte dal Coordinamento generale statistico attuariale dell'Istituto. Ciò non elimina la necessità di ulteriori verifiche. L'attività di Inapp si concentra sulla validazione della coerenza interna delle informazioni contributive e retributive, sull'armonizzazione delle variabili con le altre fonti e sulla ricostruzione coerente delle traiettorie lavorative individuali nel tempo. La presenza di record annuali distinti per impresa e caratteristiche contributive del lavoratore richiede infatti un attento lavoro di ricomposizione longitudinale, volto a garantire continuità e correttezza nell'osservazione dei percorsi occupazionali. Le basi dati Istat, caratterizzate da un elevato grado di standardizzazione, richiedono prevalentemente operazioni di allineamento classificatorio e di coerenza rispetto al perimetro di osservazione adottato nelle altre fonti. Al termine di tali operazioni vengono prodotti archivi standardizzati, controllati e pseudonimizzati, idonei all'integrazione micro tra le diverse fonti informative. Questa fase preliminare è condizione necessaria per la successiva attività di *record linkage*, dalla quale deriva la costruzione della base dati *employer-employee* utilizzata nell'analisi empirica di questo studio.

Record linkage e integrazione micro tra archivi

La fase di integrazione tra le diverse fonti informative non si esaurisce in un semplice processo tecnico di accostamento tra archivi, ma costituisce un passaggio metodologicamente complesso che richiede la definizione preliminare dell'universo di riferimento, delle chiavi di collegamento e dei criteri di validazione degli abbinamenti effettuati. Nel caso in esame, l'integrazione avviene prevalentemente mediante *linkage* deterministico basato sugli identificativi fiscali delle imprese e dei lavoratori. L'utilizzo del codice fiscale come chiave di collegamento garantisce, in linea teorica, un'elevata affidabilità dell'abbinamento, in quanto si tratta di

un identificativo univoco e trasversale alle diverse basi amministrative. Tuttavia, la presenza di universi solo parzialmente coincidenti impone un'attenta valutazione degli esiti del *linkage*. In primo luogo, occorre distinguere tra mancati abbinamenti 'leciti', dovuti a differenze strutturali tra gli archivi (ad esempio, imprese presenti in ASIA ma non soggette a Comunicazioni Obbligatorie in un determinato periodo, oppure lavoratori non inclusi nel campione Inps), e mancati abbinamenti imputabili a incongruenze negli identificativi o a disallineamenti temporali. La valutazione della qualità del *linkage* richiede pertanto l'analisi sistematica delle unità abbinate e non abbinate, nonché il confronto tra variabili comuni alle diverse fonti al fine di verificarne la coerenza. Nel processo di integrazione assumono particolare rilievo anche le trasformazioni societarie e le modifiche degli identificativi nel tempo, che possono incidere sia sulla ricostruzione delle carriere lavorative sia sulla corretta attribuzione dei lavoratori alle imprese osservate nella RIL. Tali aspetti vengono affrontati mediante procedure di controllo incrociato e verifiche longitudinali sugli identificativi, al fine di preservare la continuità informativa delle unità osservate. L'integrazione micro tra archivi costituisce dunque una fase critica dell'intero processo, nella quale la dimensione tecnica del record linkage si intreccia con scelte metodologiche relative alla definizione dell'universo analitico e alla qualità delle informazioni utilizzate.

Bibliografia

Abowd J.M., Kramarz F., Margolis D.N. (1999), High wage workers and high wage firms, *Econometrica*, 67, n.2, pp.251-333

Andrews D., Criscuolo C., Gal P.N. (2016), *The Best versus the Rest: The Global Productivity Slowdown, Divergence across Firms and the Role of Public Policy*, OECD Productivity Working Papers n.5, Paris, OECD Publishing

Babet D., Godechot O., Palladino M.G. (2025), In the Land of AKM: Explaining the Dynamics of Wage Inequality in France, *Journal of Labor Economics*, <DOI: 10.1086/740187>

Banca d'Italia (2025), *Relazione annuale sul 2024*, Roma, Banca d'Italia

Barth E., Bryson A., Davis J.C., Freeman R. (2016), It's where you work: Increases in earnings dispersion across establishments and individuals in the United States, *Journal of Labor Economics*, 34, n.S2, pp.S67-S97

Becker G.S. (1964), *Human capital: A theoretical and empirical analysis, with special reference to education*, New York, National Bureau of Economic Research

Bonacini L., Patriarca F., Santoni E. (2025), Background wage premia, beyond education: Firm sorting and unobserved abilities of graduates, *Economics of Education Review*, 109, Article 102733

Bonhomme S., Lamadon T., Manresa E. (2019), A distributional framework for matched *employer-employee* data, *Econometrica*, 87, n.3, pp.699-739

Bugamelli M., Lotti F., Amici M., Ciapanna E., Colonna F., D'Amuri F., Giacomelli S., Linarello A., Manaresi F., Palumbo G., Scoccianti F., Sette E. (2018), *Productivity Growth in Italy: A Tale of a Slow-Motion Change*, Questioni di Economia e Finanza (Occasional Papers) n.422, Roma, Banca d'Italia

Card D., Cardoso A.R., Heining J., Kline P. (2018), Firms and Labor Market Inequality: Evidence and Some Theory, *Journal of Labor Economics*, 36, n.S1, pp.S13-S70

Card D., Heining J., Kline P. (2013), Workplace heterogeneity and the rise of West German wage inequality, *Quarterly Journal of Economics*, 128, n.3, pp.967-1015

Casarico A., Lattanzio S. (2024), What firms do: Gender inequality in linked *employer-employee* data, *Journal of Labor Economics*, 42, n.2, pp.325-355

Devicienti F., Fanfani B., Maida A. (2019), Collective bargaining and the evolution of wage inequality in Italy, *British Journal of Industrial Relations*, 57, n.2, pp.377-407

Inps (2024), *XXIII Rapporto annuale*, Roma, Inps

Kline P., Petkova N., Williams H., Zidar O. (2019), Who profits from patents? Rent-sharing at innovative firms, *Quarterly Journal of Economics*, 134, n.3, pp.1343-1404

Lachowska M., Mas A., Saggio R., Woodbury S. (2023), Do firm effects drift? Evidence from Washington administrative data, *Journal of Econometrics*, 233, n.2, pp.375-395

Lamadon T., Mogstad M., Setzler B. (2022), Imperfect competition, compensating differentials, and rent sharing in the U.S. labor market, *American Economic Review*, 112, n.1, pp.169-212

Naticchioni P., Ricci A., Rustichelli E. (2010), Far away from a skill-biased change: falling educational wage premia in Italy, *Applied Economics*, 42, n.26, pp.3383-3400

OECD (2024), *OECD Employment Outlook 2024*, Paris, OECD Publishing

Pan W. (2025), Firm-specific pay premia and the returns to higher education: Evidence from community colleges, *Labour Economics*, 84, Article 102490

Song J., Price D.J., Guvenen F., Bloom N., von Wachter T. (2019), Firming up inequality, *Quarterly Journal of Economics*, 134, n.1, pp.1-50

Pence M. (1973), Job market signaling, *Quarterly Journal of Economics*, 87, n.3, pp.355-374

Van Reenen J. (1996), The creation and capture of rents: Wages and innovation in a panel of U.K. companies, *Quarterly Journal of Economics*, 111, n.1, pp.195-226

Irene Brunetti
i.brunetti@inapp.gov.it
Ricercatrice in economia applicata presso Inapp e responsabile scientifica di un accordo di collaborazione tra la Presidenza del Consiglio dei ministri e Inapp. I suoi interessi di ricerca riguardano principalmente le dinamiche del mercato del lavoro e la valutazione delle politiche pubbliche. Fra le pubblicazioni recenti si segnalano: Work-from-Home Job Creation as a Health-Risk Mitigation Strategy: Firm-Level Evidence from the COVID-19 Pandemic, *American Journal of Health Economics*, forthcoming, 2026; Peer Interactions, Local Markets, and Within Occupation Wages: Evidence from Italy, *Journal of Regional Science*, 2025.

Michelangelo Filippi
m.filippi@repnet.it
Ricercatore esperto della trasformazione a fini statistici dei dati amministrativi, collabora con università ed enti di ricerca pubblici e privati. È specializzato nell'elaborazione e analisi di dati micro. La sua attività di ricerca si concentra prevalentemente su vari aspetti del mercato del lavoro e della salute.

Valentina Gualtieri
v.gualtieri@inapp.gov.it
Appassionata di metodologia statistica è attualmente responsabile del Servizio statistico dell'Inapp e National Project manager del Programma OCSE-PIAAC (Programme for the International Assessment of Adult Competencies). È specializzata nella pianificazione di indagini statistiche in ambito sociale e nell'elaborazione e analisi di dati micro. La sua attività di ricerca si concentra prevalentemente sull'analisi di fenomeni legati all'investimento in istruzione, al capitale umano, al mercato del lavoro, alla disuguaglianza salariale e alla qualità del lavoro.

Andrea Ricci
an.ricci@inapp.gov.it
Dirigente di ricerca in economia ed economia applicata presso l'Inapp, dove coordina la Struttura di ricerca Imprese, transizioni e sviluppo del Mezzogiorno. I suoi principali interessi di ricerca hanno per oggetto la valutazione delle politiche per il lavoro, l'economia del mercato del lavoro, l'organizzazione industriale e il comportamento delle imprese. È autore di numerose ricerche pubblicate su riviste nazionali e internazionali.

La mobilità regionale degli studenti universitari

Opportunità e prime evidenze da una sperimentazione di linkage tra Anagrafe nazionale studenti e banca dati WHIP

Andrea Priulla
Università di Enna-Kore

Dalit Contini
Università di Torino

Sonia Della Monica
Università di Torino

Noemi Ferrara
Università di Torino

Il contributo presenta una sperimentazione di *linkage privacy-preserving* tra Anagrafe nazionale studenti e banca dati WHIP, finalizzata a integrare informazioni sui percorsi universitari e lavorativi. Il lavoro descrive la procedura di abbinamento, valuta le caratteristiche dell'archivio integrato risultante e ne illustra il potenziale informativo per lo studio della mobilità interregionale tra università e lavoro. I risultati preliminari evidenziano differenze territoriali nei percorsi dei laureati e degli studenti che interrompono gli studi.

Regional mobility of university students. Opportunities and preliminary evidence from a linkage procedure between the National student registry and the work histories Italian Panel

The paper presents a privacy-preserving linkage procedure between the National Student Registry and the WHIP database, aimed at integrating information on university and employment trajectories. It describes the matching procedure, assesses the characteristics of the resulting integrated archive, and illustrates its analytical potential for studying interregional mobility between university and work. The preliminary results highlight territorial differences in the trajectories of graduates and students who discontinue their studies.

DOI: 10.53223/Sinappsi_2026-02-7

Citazione
Priulla A., Contini D., Della Monica S., Ferrara N. (2026), La mobilità regionale degli studenti universitari. Opportunità e prime evidenze da una sperimentazione di linkage tra Anagrafe nazionale studenti e banca dati WHIP, *Sinappsi*, XVI, n.2, pp.106-121

Parole chiave
Dati amministrativi
Mobilità giovanile
Studenti

Keywords
Administrative data
Youth mobility
Students

Introduzione

La mobilità regionale degli studenti universitari e dei laureati è diventata negli ultimi anni uno dei temi centrali nel dibattito sulle disuguaglianze territoriali in Italia. Il trasferimento di giovani dalle regioni del Mezzogiorno verso i Poli metropolitani del Centro-Nord e verso l'estero per studio o per lavoro è spesso letto in chiave di 'fuga dei cervelli', con implicazioni importanti per le traiettorie di sviluppo regionale e per la progettazione delle politiche di istruzione e lavoro (Dotti *et al.* 2013; Priulla *et al.* 2025; Tosi *et al.* 2025).

Nonostante la rilevanza del tema, l'evidenza empirica disponibile sulla mobilità regionale dopo la laurea rimane ancora parziale. Attualmente, la principale fonte informativa è rappresentata dalle indagini AlmaLaurea, che combinano archivi amministrativi universitari con interviste rivolte all'intera popolazione dei laureati degli atenei che aderiscono al Consorzio, al momento della laurea e a 1, 3 e 5 anni dal conseguimento del titolo. Queste indagini offrono informazioni molto dettagliate sui corsi di studio, sulle aspirazioni e sulle prime esperienze di lavoro, e sono ampiamente utilizzate

per il monitoraggio e la valutazione delle politiche universitarie.

Tuttavia, i dati AlmaLaurea presentano alcuni limiti che rendono difficile l'analisi delle biografie di vita individuali. Anzitutto, l'orizzonte temporale delle informazioni è limitato a specifici momenti della vita degli individui, senza la possibilità di ricostruire le sequenze di episodi lavorativi, disoccupazione, inattività e gli eventuali spostamenti di regione nel medio periodo. In secondo luogo, si pone un problema di *attrition*, ovvero di riduzione del tasso di risposta alle indagini successive alla laurea, che diminuisce nel tempo e in modo non casuale: tendono infatti a uscire dal campione soprattutto i laureati triennali, introducendo potenziali distorsioni nelle stime di mobilità e degli esiti occupazionali. Infine, i dati AlmaLaurea consentono di osservare esclusivamente gli individui che hanno conseguito la laurea, mentre non raccolgono informazioni su chi non ha concluso il percorso universitario, perché ancora iscritto o perché ha abbandonato gli studi.

Accanto a questa fonte, l'Italia dispone di due grandi famiglie di archivi amministrativi potenzialmente cruciali per lo studio congiunto di istruzione terziaria e lavoro: l'Anagrafe nazionale degli studenti (ANS), che segue longitudinalmente le carriere di studio degli individui iscritti a tutte le università italiane e la banca dati WHIP (*Work and Health Italian Panel*), che ricostruisce – a partire dagli archivi Inps, Inail, del Ministero della Salute e del Ministero del Lavoro – le biografie di lavoro e salute di un campione rappresentativo di lavoratori in Italia, includendo in particolare le Comunicazioni Obbligatorie (COB) su attivazioni, trasformazioni e cessazioni dei contratti di lavoro, che contengono informazioni molto dettagliate sulla professione svolta.

Se i dati ANS potessero essere collegati a dati amministrativi di fonte Inps o Ministero del Lavoro (quali WHIP e COB), sarebbe possibile seguire gli individui dal luogo di nascita e di scuola superiore, al luogo di iscrizione universitaria, fino ai luoghi di lavoro effettivi, ricostruendo così percorsi di mobilità regionale in una prospettiva longitudinale. Tuttavia, gli archivi sono gestiti da soggetti diversi, sono stati costruiti per finalità amministrative distinte e attualmente non sono pensati per essere integrati a fini di ricerca. Pertanto, salvo alcuni esperimenti circoscritti a singoli atenei o regioni, mancano infrastrutture dati stabili che consentano di usare congiuntamente queste fonti su scala nazionale (Leombruni e Ardito 2026).

Questo articolo presenta una sperimentazione di integrazione tra l'archivio ANS e WHIP, con la finalità di esplorarne l'utilità in un'ottica di *policy analysis*. In particolare, ci si prefigge di: a) mostrare come, con una procedura di *linkage privacy-preserving*, sia possibile mettere in relazione i due archivi amministrativi; b) discutere il potenziale di questa infrastruttura per lo studio della mobilità interna interregionale degli studenti e dei laureati nelle transizioni tra università e lavoro.

1. Quali archivi per quali potenzialità

La diffusione di dati amministrativi in ambito di ricerca economica e sociale è stata accompagnata da una crescita di studi che ne sfruttano la copertura di popolazione, la profondità temporale e la relativa economicità rispetto alle indagini campionarie. In Italia, diversi lavori impiegano i dati derivanti dall'Anagrafe nazionale studenti (ANS), per studiare la mobilità di studio e le carriere degli studenti universitari (D'Agostino *et al.* 2019; Attanasio *et al.* 2020; Usala *et al.* 2023; Priulla *et al.* 2025); altri lavori utilizzano i dati di fonte Inps per approfondire molteplici aspetti del mercato del lavoro (Berton *et al.* 2012; Galizzi *et al.* 2023; Devicienti *et al.* 2025; Ardito *et al.* 2026 solo per fare alcuni esempi). Al tempo stesso, la riflessione metodologica su come questi dati vengono generati, integrati e interpretati è ancora meno strutturata rispetto a quella maturata attorno alle indagini statistiche tradizionali (Leombruni 2024).

Ad oggi, infatti, la principale fonte dati utilizzata per esplorare il percorso universitario e lavorativo dei laureati italiani è l'indagine AlmaLaurea, che offre una base informativa sulle caratteristiche dei corsi, delle carriere universitarie e dei primi esiti occupazionali. Tuttavia, come discusso nella sezione precedente, oltre a riguardare esclusivamente coloro che conseguono il titolo universitario, i dati AlmaLaurea presentano diversi limiti che rendono difficile affrontare domande di ricerca in una prospettiva di biografia di vita, ovvero seguire nel tempo le traiettorie di studio e lavoro degli individui.

In particolare, l'indagine raccoglie informazioni (autoriportate) in specifici momenti nel tempo – al momento della laurea e successivamente a 1, 3 e 5 anni – senza permettere una ricostruzione continua dei percorsi individuali. A questo si aggiunge un problema di *attrition* non casuale che riduce, e presumibilmente distorce, l'informazione disponibile sulle traiettorie degli individui: le indagini successive a quella del momento della laurea presentano infatti tassi di risposta progressivamente più bassi, con una caduta particolarmente marcata tra i laureati triennali (Appendice, tabella A1). Quest'ultima è legata principalmente al disegno dell'indagine: nel questionario somministrato a 1, 3 e 5 anni dalla laurea triennale viene chiesto se il rispondente si sia iscritto a un corso magistrale. In caso di risposta affermativa, il questionario si interrompe e l'intervistato viene contattato nuovamente solo al

conseguimento del titolo magistrale. Non è quindi possibile osservare eventuali episodi di lavoro che si verificano durante il percorso di laurea magistrale e se lo studente abbandona gli studi nel corso della magistrale, viene perso dall'osservazione.

In aggiunta alla difficoltà di ricostruire le biografie di lavoro, un ulteriore limite dei dati AlmaLaurea è rappresentato dalla natura delle informazioni sulla professione, che nelle interviste a 1 e 3 anni seguono una classificazione ad hoc e con un grado di dettaglio limitato, mentre solamente nell'intervista a 5 anni si fa riferimento alla codifica Istat a 5 digit.

In questo contesto, gli archivi amministrativi longitudinali di fonte Inps e Ministero del Lavoro potrebbero rappresentare una risorsa preziosa per seguire intere coorti di individui nel tempo, osservando il percorso di studio – tramite le informazioni su regione di frequenza, tipologia del diploma, percorso accademico inclusi eventuali cambi di corso di laurea e abbandoni – e i vari episodi di lavoro, con indicazioni puntuali sulla professione svolta, nonché eventuali periodi di disoccupazione e episodi di mobilità territoriale.

Il problema, tuttavia, è che attualmente gli archivi amministrativi relativi a università e lavoro non sono integrati in modo sistematico. L'ANS fornisce informazioni dettagliate sui percorsi accademici, ma non dice nulla sulla carriera occupazionale. Allo stesso tempo WHIP e COB descrivono le carriere lavorative, ma includono la sola informazione sul massimo titolo di studio conseguito, che è tuttavia incompleta e di qualità molto incerta; inoltre, non vengono registrati né il luogo di conseguimento del titolo né il corso di laurea o l'area disciplinare.

In questo contesto, la sperimentazione si propone come un caso concreto di integrazione di fonti amministrative, basato su una procedura di *linkage privacy-preserving* che consente di mettere in relazione archivi di natura diversa. L'obiettivo non è soltanto sfruttare dati già raccolti, ma mostrare come la loro integrazione possa consentire di rispondere a nuovi interrogativi di ricerca. In particolare, esploriamo qui su un tema – la mobilità interregionale degli studenti (laureati e non) tra università e lavoro – che potrebbe beneficiare sostanzialmente dall'analisi di dati amministrativi longitudinali.

L'Anagrafe nazionale degli studenti

Le fonti di dati impiegate dal nostro gruppo di ricerca per l'esecuzione di questa sperimentazione sono l'Anagrafe nazionale degli studenti e WHIP. L'ANS è l'archivio amministrativo che registra le carriere di tutti gli studenti iscritti alle università italiane, fornendo informazioni anche sul percorso scolastico pregresso, quali la tipologia di diploma, nonché l'anno di conseguimento e l'area geografica dell'istituto frequentato. Va tuttavia segnalata una limitazione dei dati dell'ANS, rappresentata dall'assenza di informazioni sugli individui che si iscrivono in università all'estero, o che si trasferiscono all'estero dopo l'immatricolazione nel sistema universitario italiano.

Per ciascun individuo sono disponibili i seguenti dati:

- caratteristiche anagrafiche (sesso, data, Comune e regione di nascita, cittadinanza ecc.);
- informazioni sulla scuola superiore (tipo di diploma, anno di conseguimento, area geografica dell'istituto, voto di diploma ecc.);
- dettagli a cadenza annuale sulle carriere universitarie (università, corso di studio, tipo di laurea, data di immatricolazione, crediti formativi universitari (CFU), eventuali passaggi di corso o ateneo, data di conseguimento della laurea, interruzioni).

I dati dell'ANS sono stati resi disponibili grazie alla stipula di un protocollo di ricerca tra il MUR, Invalsi e alcune università con capofila l'Università di Palermo, nell'ambito di un progetto PRIN finalizzato allo studio della mobilità di studio dei giovani italiani[1]. L'utilizzo di questi dati (confluiti in un archivio denominato Mobysu.it, prodotto nel 2017 e in seguito aggiornato periodicamente) ha consentito di superare le limitazioni degli archivi dei singoli atenei, che non permettono di seguire gli studenti in caso di trasferimento tra università.

Per la sperimentazione di *linkage* sono stati impiegati i dati relativi alle coorti di immatricolati in tutte le università italiane negli anni accademici dal 2009/10 al 2016/17 (con l'esclusione dell'anno 2011 dovuta a problemi di dati mancanti sul Comune di nascita degli studenti). Ciascuna coorte viene osservata fino al 30 giugno 2019. Nello specifico, la sottopopolazione dei dati dell'ANS utilizzata è costituita da 96.968 individui, derivata dall'applicazione dello stesso schema di campionamento impiegato per WHIP, ovvero uno schema di campionamento sistematico basato su 2 date di nascita per ogni mese, a seguito della quale

1 I dati utilizzati in questo studio sono stati elaborati in conformità al protocollo di ricerca per lo studio *From high school to the job placement: analysis of university careers and university mobility from Southern to Northern Italy* tra il Ministero dell'Università e della Ricerca, il Ministero dell'Istruzione e del merito, l'Università degli Studi di Palermo come ente capofila, e l'Istituto Invalsi. Sono coinvolte inoltre le università di Cagliari, Firenze, Napoli Federico II, Sassari, Siena, Torino (in seguito anche Enna KORE, Salerno, Bologna).

sono state effettuate diverse operazioni di selezione. In primo luogo, sono stati considerati coloro la cui età al momento della prima immatricolazione risultava non superiore a 22 anni. Sono stati esclusi anche tutti gli studenti che, in un qualsiasi momento della propria carriera universitaria siano risultati iscritti ad un'università telematica. Questa scelta risiede nel fatto che per frequentare un corso di laurea offerto da un'università telematica non è necessario trasferirsi nella regione sede dell'ateneo. Infine, è stato rimosso un numero limitato di record che presentavano anomalie strutturali nelle carriere (duplicazioni di ID, sequenze di immatricolazioni e lauree logicamente impossibili, sovrapposizioni non risolvibili tra corsi diversi), per cui non era possibile ricostruire in modo credibile il percorso di studio. Per concludere, sono stati esclusi gli studenti immatricolati ad un corso di laurea a ciclo unico, poiché costituiscono percorsi al cui termine si aspira generalmente ad un percorso lavorativo che richiede l'iscrizione ad un albo o ad una professione normata, non incluse all'interno del database WHIP.

WHIP e l'archivio delle Comunicazioni Obbligatorie

Il secondo archivio di dati amministrativi utilizzato per la sperimentazione di *linkage* è WHIP, una banca dati che ricostruisce le biografie lavorative e di salute di un campione di circa il 7% dei lavoratori italiani (tramite l'estrazione su due date mensili di nascita), il cui titolare è il Ministero della Salute ed è costruita a partire da diversi archivi amministrativi[2]. Le fonti che compongono questo database sono: l'Inps con dati sugli episodi lavorativi relativi alla maggior parte delle forme di lavoro in Italia (dipendenti, parasubordinati, professionisti, artigiani e commercianti), nonché le dichiarazioni ISEE e le informazioni su benefici di welfare (quali pensioni o indennità di disoccupazione) e il Ministero del Lavoro con le Comunicazioni Obbligatorie (COB), che forniscono informazioni dettagliate sui contratti di lavoro subordinato e parasubordinato pubblico e privato dal 2008 al 2019 (ultimo aggiornamento a nostra disposizione)[3].

Complessivamente, nei dati WHIP sono comprese:

- tutte le fattispecie di lavoro registrate nelle COB, incluse collaborazioni non occasionali, specializzandi e dottorandi;
- Partite IVA (artigiani, professionisti, commercianti non con cassa separata).

Non sono compresi invece albi a gestione separata e alcune forme di lavoro occasionale.

Date le finalità dello studio, tra tutte le informazioni presenti nella banca dati WHIP sono state considerate esclusivamente le variabili anagrafiche di base (sesso, data di nascita, Comune di nascita) e le principali informazioni sui rapporti di lavoro (inizio e fine contratto, regione di lavoro, classificazione della professione). Rispetto al campione di individui selezionati per il *linkage*, sono stati considerati solo individui appartenenti alle stesse coorti selezionate dal lato ANS, nati in Italia e che abbiano presentato almeno una dichiarazione ISEE. Le ultime due selezioni sono motivate dal fatto che nella chiave di abbinamento è incluso il Comune di nascita, informazione disponibile solo per gli individui nati in Italia ed esclusivamente sul flusso informativo ISEE[4].

2. Il disegno di linkage

Nel seguito utilizziamo il termine *linkage* per indicare la procedura complessiva di integrazione tra il dataset Mobysu.it di fonte ANS e il dataset WHIP. Con il termine *match* indichiamo invece l'esito della procedura, ossia una possibile corrispondenza tra uno o più record dei due archivi. Come sarà chiarito nei paragrafi successivi, la procedura adottata in questa fase si basa su una chiave di abbinamento che *non* consente l'identificazione certa dei soggetti (sesso, Comune e data di nascita).

Entrambi gli archivi, ancorché privi di identificativi diretti, contengono variabili che – soprattutto se combinate tra loro – possono aumentare il rischio di reidentificazione: può infatti accadere che un individuo presenti una combinazione unica di caratteristiche che ne consente una potenziale identificazione indiretta. Nonostante in questo studio si consideri una chiave di abbinamento basata su un insieme minimale di informazioni, è stato comunque adottato, per le procedure di *linkage*, un disegno a doppio cieco, costituito da due gruppi di lavoro nettamente distinti per compiti e informazioni accessibili.

La distinzione tra gruppi di lavoro risponde a un principio di minimizzazione delle informazioni

2 Vedi https://www.epi.piemonte.it/whipsalute/.

3 WHIP include anche dati sugli infortuni sul lavoro e le malattie professionali (Inail), e le Schede di dimissione ospedaliera rilasciati dal Ministero della Salute.

4 Ha presentato una dichiarazione ISEE il 46% dei lavoratori delle coorti di nascita di interesse presenti nel file WHIP. Tuttavia, è probabile che tra gli studenti universitari questa quota sia sensibilmente più elevata, date le forti riduzioni nelle tasse universitarie previste per i redditi medi e bassi. Nel complesso, è plausibile che il campione analizzato sia maggiormente rappresentativo di una popolazione con livelli di reddito non elevati.

accessibili a ciascuno: chi effettua il *linkage* iniziale non dispone delle informazioni sostantive sulle carriere, mentre chi analizza le biografie non accede alle variabili anagrafiche disaggregate utilizzate per costruire la chiave di *linkage*.

Il Team A ha operato su due file, uno derivante dal database ANS e l'altro da WHIP, senza però sapere quale dei due provenisse da quale archivio, ognuno composto esclusivamente da:

- un identificativo interno anonimo;
- sesso;
- data di nascita completa (giorno, mese, anno);
- Comune di nascita.

Su questi due file il Team A ha applicato una regola di *linkage* basata sull'esatta coincidenza di sesso, data di nascita e Comune di nascita, identificando tutti i possibili gruppi di record che condividevano la stessa chiave. Il dataset risultante è quindi composto dalle diverse possibili configurazioni. Match univoci (match 1:1): un record ANS e un record WHIP; match non univoci: un record ANS associato a più record WHIP (1:molti); più record ANS associati a un unico record WHIP (molti:1); più record ANS collegati a più record WHIP (molti:molti). Inoltre vi sono record ANS che non vengono abbinati ad alcun record WHIP.

Poiché Comune e data di nascita sono le informazioni che contribuiscono maggiormente al rischio di reidentificazione, ma non sono necessarie all'analisi, il Team A restituisce al Team B un file in cui tali informazioni sono riportate a un livello maggiore di aggregazione. In particolare, il file contiene: identificativi anonimi, sesso, regione e anno di nascita.

Su questo file, il Team B associa, per ciascun identificativo, un set di variabili longitudinali:

- per i record provenienti da ANS: informazioni sintetiche su scuola superiore, università, tipo di corso, data di immatricolazione e di laurea, eventuali cambi di corso/ateneo;
- per i record provenienti da WHIP: informazioni sui rapporti di lavoro (periodo di inizio/fine, tipologia contrattuale, regioni di lavoro, descrizione della professione).

In questo modo, il Team B dispone di un insieme di biografie di studio e lavoro organizzate in diversi gruppi di match, senza tuttavia disporre delle variabili chiave nella loro forma disaggregata. Ciò consente di tutelare la privacy e, al tempo stesso, di valutare il potenziale informativo dei dati.

Entriamo ora nel merito dei dati così prodotti.

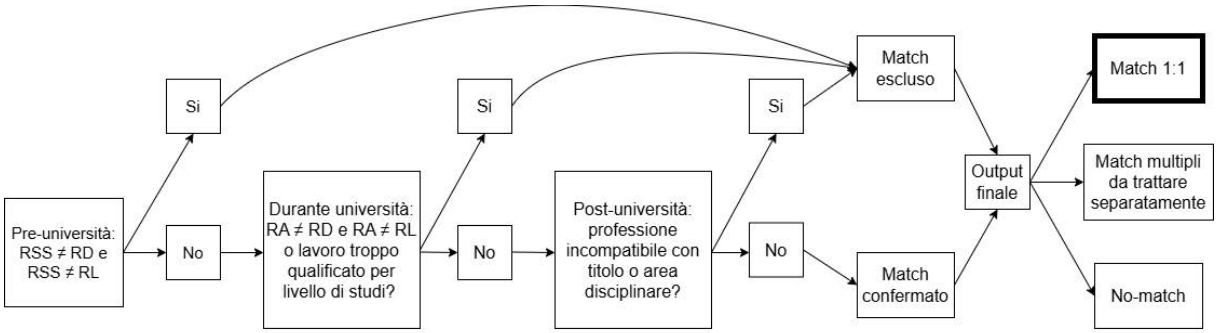
1) In primo luogo, come si può interpretare l'assenza di un match valido? Lo scenario più plausibile è che l'individuo non abbia avuto esperienze lavorative. Tuttavia, non si possono escludere altre casistiche: ad esempio, l'individuo potrebbe lavorare all'estero, svolgere un'attività non ricompresa tra le fattispecie registrate in WHIP, oppure non avere mai presentato una dichiarazione ISEE (in questo caso, come descritto sopra, l'informazione sul Comune di nascita, che costituisce parte della chiave di *linkage*, non è disponibile dal lato WHIP).
2) Nel caso di match 1:1, l'abbinamento tra un solo studente – identificato da una combinazione univoca di sesso, data di nascita e Comune di nascita – e un solo lavoratore non garantisce che la biografia individuale sia ricostruita con certezza. Nello stesso giorno e luogo potrebbe infatti essere nato un individuo con esperienza lavorativa ma che non ha frequentato l'università (e quindi non risulta nell'archivio ANS), mentre lo studente potrebbe aver lavorato all'estero o non essere ancora entrato nel mercato del lavoro (e quindi non compare in WHIP). Ciò comporta che, sebbene la procedura di *linkage* sia basata su regole esatte di corrispondenza tra chiavi di aggancio predefinite, nessun match sia configurabile come risultato di un abbinamento certo, perché più persone possono condividere la stessa combinazione di sesso, data e Comune di nascita. La probabilità che i due record abbinati univocamente si riferiscano alla stessa persona decresce all'aumentare della dimensione del Comune, perché diventa via via più plausibile che vi siano più persone nata nella stessa data[5].
3) Anche nei match 1:molti o molti:1 – che, come evidenziato nelle conclusioni, costituiranno l'oggetto di un proseguimento di questo lavoro tramite abbinamento probabilistico – non è certo che tra i record presenti nel lato 'molti' vi sia effettivamente quello riferito alla stessa persona presente nel lato singolo.

3. Regole di coerenza biografica per l'affinamento dei match

L'affidabilità dei match può essere ulteriormente migliorata mediante l'applicazione di alcune regole di esclusione e di coerenza biografica, organizzate

5 Condizionatamente all'osservazione di un match 1:1, per i laureati la probabilità che i due record si riferiscano allo stesso individuo è pari a circa il 98% in un Comune di 10.000 abitanti e al 70% in un Comune di 100.000 abitanti. Tale stima è stata ottenuta assumendo un tasso di natalità pari a quello medio italiano del periodo, insieme a ipotesi semplificatrici sulla probabilità di conseguire una laurea e di essere occupati tra i laureati. I risultati variano solo marginalmente al variare di tali ipotesi.

Figura 1. Flow-chart delle regole di affinamento applicate alle coppie di match studente-lavoratore generati dalla procedura di linkage



RSS = Regione scuola superiore; RD = Regione domicilio; RL = Regione lavoro; RA = Regione ateneo.

Fonte: elaborazione degli Autori

lungo l'asse temporale delle carriere di studio (prima dell'università, durante l'università, dopo l'università) e integrate con informazioni ausiliarie sui percorsi lavorativi, provenienti dalle COB. Le regole di esclusione hanno lo scopo di identificare coppie di record associate con informazioni incoerenti tra loro, articolate in tre fasi temporali: 1) precedente l'immatricolazione ad un corso di laurea; 2) durante il periodo universitario; 3) successiva al conseguimento di un titolo di laurea, triennale o magistrale.

Uno schema sintetico dei passaggi effettuati è visibile in figura 1.

Nella fase pre-universitaria utilizziamo l'informazione, presente in ANS, relativa alla regione in cui è stato conseguito il diploma di scuola superiore (RSS) e la confrontiamo con le informazioni provenienti da WHIP sulla regione di domicilio (RD) e sulla regione in cui si sono svolti gli eventuali episodi di lavoro precoci (RL). Un match viene escluso quando la regione della scuola superiore è diversa sia dalla regione di domicilio sia dalla regione di lavoro, ossia nei casi in cui valga contemporaneamente RSS ≠ RD e RSS ≠ RL. La logica sottostante è la seguente: dato uno studente x e un lavoratore y che condividono la stessa chiave anagrafica di *linkage*, è più probabile che siano la stessa persona se, nel periodo pre-universitario, la regione in cui x frequenta le superiori coincide con la regione in cui y vive e lavora, dal momento che l'eventuale lavoro viene svolto in contemporanea con gli studi. Quando questa condizione di coerenza territoriale non è soddisfatta, la coppia viene scartata perché ritenuta poco plausibile.

Nella fase relativa agli anni di iscrizione ad un corso di laurea, applichiamo una regola analoga, sostituendo alla regione della scuola superiore la regione dell'ateneo (RA). Vengono esclusi i match in cui la regione dell'università è diversa sia dalla regione di domicilio sia dalla regione di lavoro, cioè i casi in cui RA ≠ RD e RA ≠ RL. La logica è molto simile a quella della regola precedente: se studio, residenza e (eventuale) lavoro si collocano per l'intero periodo in regioni completamente diverse, senza alcuna sovrapposizione territoriale, il match è ritenuto non plausibile.

A questa regola generale si aggiunge un controllo specifico sulla coerenza fra livello di studi e tipo di occupazione: vengono esclusi gli studenti che hanno abbandonato il percorso universitario e che risultano aver svolto un lavoro per il quale è richiesta una laurea, senza aver conseguito almeno un titolo triennale (con eccezione degli studenti vicini alla laurea, che potrebbero essere stati assunti come laureati perché in procinto di ottenerla). Nel complesso, l'applicazione delle regole per la fase universitaria e pre-universitaria seleziona un sottoinsieme di match 1:1 che mantiene la coerenza territoriale fra studi, domicilio e lavoro e la compatibilità fra livello di studi effettivamente raggiunto e qualificazione del lavoro svolto.

La fase post-universitaria si concentra sulla coerenza tra titolo di studio e occupazione dopo l'uscita dall'università. Le esclusioni si basano su criteri più puntuali di congruenza tra laurea e tipo di lavoro, con particolare attenzione alle professioni regolamentate o per le quali è richiesta una specifica formazione universitaria. L'idea di fondo è che non sia plausibile che un individuo presente in ANS senza il titolo richiesto coincida con un lavoratore WHIP occupato in una professione per cui quel titolo è obbligatorio: ad esempio, chi lavora come veterinario deve avere la laurea in medicina veterinaria; in caso contrario, è verosimile che si tratti di due persone diverse.

Operativamente, viene applicata una serie di filtri. Si eliminano innanzitutto i casi in cui compaiono lavori in area sanitaria senza una laurea medica. Vengono poi esclusi i docenti privi del titolo adeguato e i laureati triennali che risultano docenti universitari. Un controllo specifico riguarda gli studenti iscritti a corsi triennali che risultano contemporaneamente ricercatori o docenti, così come i non laureati occupati in posizioni accademiche (ricercatori, docenti o altre figure universitarie).

Ulteriori filtri riguardano la coerenza tra area disciplinare e professioni ad alta specializzazione tecnica: vengono ad esempio esclusi i laureati triennali in ambito umanistico che risultano lavorare come ingegneri per periodi superiori a 30 giorni. Un ulteriore filtro elimina i casi di dottorato senza laurea: in queste situazioni, qualora la laurea fosse stata conseguita all'estero, l'individuo non sarebbe comunque osservabile in ANS.

Infine, si applicano controlli sulle professioni per le quali è richiesta l'iscrizione ad albi (ingegneri, commercialisti e altre professioni ordinistiche), escludendo le coppie in cui il lavoratore svolge un'attività soggetta ad albo ma non possiede in ANS una laurea coerente.

A seguito di queste operazioni, l'insieme iniziale di 96.968 studenti si suddivide come segue: 16.777 persone in match univoci 1:1 (17,3%), 10.117 in match 1:molti (10,4%), 16.252 in match molti:1 (15,9%), 15.453 in match molti:molti (15,9%) e 38.369 studenti senza match (39,6%).

4. Caratteristiche del dataset 1:1

I gruppi di studenti in match non univoci saranno oggetto di analisi approfondita in un prossimo futuro, per verificare se e come sia possibile assegnare a ciascuno studente una o più biografie lavorative plausibili mediante l'adozione di un approccio probabilistico. In questo lavoro ci si concentra invece sulla adeguatezza dell'utilizzo del dataset dei match 1:1 per specifiche finalità di ricerca.

Come evidenziato sopra, in questo archivio la distribuzione degli individui per dimensione del Comune di nascita è sbilanciata verso i nati in Comuni di piccola dimensione. In tabella 1 è mostrata la distribuzione degli studenti che risultano nel gruppo 1:1, nel gruppo di match non univoci e tra i no-match, per dimensione del Comune. In linea con le aspettative, gli studenti nati in Comuni di piccole dimensioni risultano sovrarappresentati tra i 1:1 e i no match rispetto a quelli nati nei grandi centri, poiché in questi Comuni la chiave utilizzata per il *linkage* è più discriminante. Più di un terzo dei match 1:1 riguarda infatti individui nati in Comuni fino ai 25.000 abitanti, mentre tra i match non univoci questa percentuale è pari al 7% circa.

Si effettua quindi un'analisi esplorativa per valutare la rappresentatività del campione 1:1 rispetto alla popolazione di riferimento, sulla base di alcune caratteristiche tipicamente associate alla carriera lavorativa. Idealmente, si vorrebbero confrontare gli studenti nel gruppo 1:1 con gli altri studenti con esperienze di lavoro. L'identificazione

Tabella 1. Distribuzione degli studenti per dimensione del Comune di nascita nei diversi gruppi di match

	Match univoci (1:1)		Match non univoci		No match	
Dimensione del Comune	N	%	N	%	N	%
< 10 mila	1.815	10,8	699	1,7	3.950	10,3
10 – 25 mila	3.875	23,1	2.247	5,4	8.523	22,2
25 – 50 mila	4.906	29,2	4.298	10,3	10.448	27,2
50 – 100 mila	3.374	20,1	5.511	13,2	7.482	19,5
100 – 250 mila	1.664	9,9	6.222	14,9	4.410	11,5
250 – 500 mila	290	1,7	3.906	9,3	1.074	2,8
500 mila – 3 milioni	72	0,4	17.214	41,2	889	2,3
NA	781	4,7	1.725	4,1	1.593	4,2
TOT	16.777	100,0	41.822	100,0	38.369	100,0

Fonte: elaborazione degli Autori

Tabella 2. Distribuzione del percorso universitario nei diversi gruppi di match

	Match univoci (1:1)		Match non univoci	
Stato del percorso	N	%	N	%
Laurea Triennale in corso	4.416	26,3	11.653	27,9
Laureato Triennale	3.178	18,9	8.293	19,8
Laurea Magistrale in corso	2.319	13,8	5.609	13,4
Laureato Magistrale	2.639	15,7	5.828	13,9
Abbandono di Laurea Magistrale	234	1,4	624	1,5
Abbandono di Laurea Triennale	3.879	23,1	9.489	22,7
Passaggio a corso a Ciclo Unico	112	0,7	326	0,8

Fonte: elaborazione degli Autori

Tabella 3. Distribuzione dell'area disciplinare nei diversi gruppi di match

	Match univoci (1:1)		Match non univoci	
Area disciplinare di laurea	N	%	N	%
Benessere e Salute	1.407	9,2	4.001	10,6
Economico-Giuridica	1.599	10,4	4.213	11,1
Formazione	429	2,8	867	2,3
Ingegneria e Architettura	1.556	10,1	3.568	9,4
Politico-Sociale	757	4,9	2.064	5,5
Scienze e Tecnologia	1.371	8,9	2.784	7,4
Umanistico-Linguistica	1.448	9,4	3.346	8,8
Non laureati	8.210	53,4	20.979	55,5

Fonte: elaborazione degli Autori

certa di quest'ultimo gruppo non è tuttavia possibile, perché la procedura di abbinamento effettuata non garantisce abbinamenti esatti. Come prima approssimazione, confrontiamo il gruppo 1:1 con l'insieme degli studenti che presentano almeno un match. Escludiamo quindi i no-match, che è plausibile che non abbiano avuto esperienze lavorative.

Le tabelle 2 e 3 mostrano le distribuzioni relative allo stato del percorso universitario (laurea triennale o magistrale, studente in corso, abbandono del percorso di studi) e all'area disciplinare del corso di laurea. Si rileva una lieve sovrarappresentazione di laureati magistrali nel gruppo 1:1, e di alcune aree disciplinari come Ingegneria e Scienze e Tecnologia.

Per analizzare se il sottoinsieme dei match 1:1 differisca sistematicamente dal resto degli individui abbinati con un approccio multivariato, sono stati stimati modelli di regressione logistica sulla probabilità di appartenere al gruppo 1:1. La variabile dipendente è un indicatore binario che assume valore 1 se l'individuo (lato ANS) presenta un match 1:1 e valore 0 se ricade in un gruppo di abbinamento non univoco. L'analisi viene dunque condotta sugli studenti che presentano almeno un match nel lato WHIP. Come variabili esplicative sono state considerate: la coorte di immatricolazione, il genere, il tipo di diploma di scuola media superiore, l'area disciplinare, lo stato del percorso universitario e la dimensione del Comune.

I risultati, riportati in appendice (tabella A2), mostrano che – con l'eccezione della dimensione del Comune di nascita[6] – la maggior parte delle variabili non presenta effetti statisticamente significativi sulla probabilità di appartenere al gruppo 1:1, oppure presenta effetti significativi ma di entità molto contenuta. Inoltre, si osserva una

6 Essere nati in Comuni più grandi riduce la probabilità di appartenere al gruppo 1:1 di diverse decine di punti percentuali rispetto alla categoria di riferimento (meno di 10.000 abitanti).

modesta differenza di genere: gli studenti maschi hanno una probabilità leggermente più elevata di appartenere al gruppo 1:1 (circa 3 punti percentuali in più rispetto alle studentesse), in linea con una maggiore presenza maschile in alcune professioni regolamentate, incluse nelle regole di esclusione riguardanti la fase post-universitaria. *Ceteris paribus*, si osserva una probabilità di appartenenza al gruppo 1:1 lievemente più elevata per le aree disciplinari Formazione e Scienze e tecnologie, mentre non si confermano le differenze relative al diverso percorso universitario.

La conclusione preliminare che possiamo trarre è che il *linkage* 1:1 non individui un gruppo di studenti molto diverso dalla popolazione di partenza, nonostante le selezioni introdotte dalla procedura utilizzata e dalla struttura degli archivi amministrativi. Ulteriori analisi dovranno essere condotte per verificare la rappresentatività del campione in termini di storie di lavoro.

È opportuno, tuttavia, ricordare che alcuni meccanismi di selezione sono strutturali al disegno del *linkage* e alla natura stessa degli archivi utilizzati. In particolare, il *linkage* privilegia implicitamente le persone che hanno avuto rapporti di lavoro regolari in settori coperti da WHIP e, dunque, effettivamente visibili nelle Comunicazioni Obbligatorie. Inoltre, risultano abbinati solo gli individui che hanno presentato almeno una dichiarazione ISEE nel periodo analizzato, perché nel lato WHIP le informazioni su data e comune di nascita sono presenti solo per questi. Tali elementi non sembrano compromettere la somiglianza complessiva tra popolazione generale e gruppo dei match 1:1, ma va ricordato che il campione ottenuto non è pienamente rappresentativo. Di conseguenza, questi meccanismi di selezione devono essere tenuti esplicitamente in conto nelle analisi empiriche attraverso analisi di sensitività e un'interpretazione dei risultati sempre ancorata ai vincoli di copertura propri dei dati amministrativi utilizzati.

5. Potenziale informativo per lo studio della mobilità interna tra università e lavoro

In questa sezione si intende fornire un'analisi preliminare della mobilità territoriale degli studenti universitari e dei laureati italiani. I risultati presentati hanno carattere illustrativo e sono finalizzati a illustrare le potenzialità del dataset costruito con la procedura di *linkage* descritta nelle sezioni precedenti.

Nel complesso, la combinazione di analisi descrittive e modelli suggerisce che il sottoinsieme dei match 1:1, pur segnato da selezioni strutturali legate alla dimensione del Comune di nascita, non presenta forti distorsioni rispetto alle principali caratteristiche degli studenti. Ciò permette di utilizzare il dataset 1:1 come base per analisi sulla mobilità regionale e sulle transizioni studio-lavoro, a condizione di tenere esplicitamente conto, nelle interpretazioni e nelle eventuali ponderazioni, dei meccanismi di selezione legati alla geografia demografica e alla copertura degli archivi amministrativi.

Grazie alla disponibilità di informazioni sulla regione di nascita, regione della scuola superiore, regione dell'ateneo e tipo/area di laurea, nonché delle sequenze di episodi lavorativi (con regioni di lavoro, settori, gruppi professionali), diventa possibile esplorare in modo più approfondito un tema centrale quale la mobilità degli studenti e dei laureati italiani, prima e dopo il conseguimento di un titolo, e i relativi percorsi professionali, offrendo una lettura più ricca dei processi di selezione territoriale rispetto a quanto consentito dalle sole indagini campionarie.

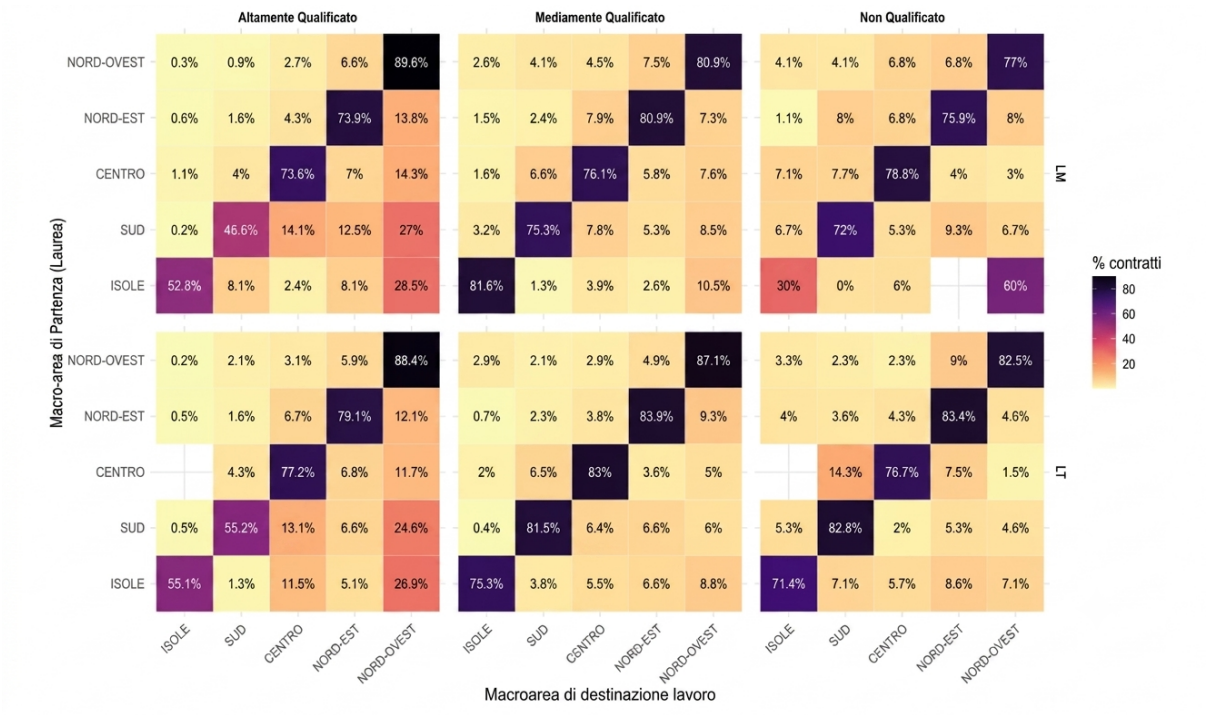
Inoltre, poiché nel dataset ANS sono presenti sia laureati sia studenti che hanno abbandonato il percorso universitario, è possibile confrontare le loro traiettorie, interrogandosi, ad esempio, se i non laureati entrino più precocemente nel mercato del lavoro ma con percorsi più localizzati e meno mobili; o se i laureati risultino più mobili territorialmente nella ricerca di un'occupazione; o ancora se vi siano differenze di mobilità e di qualità dell'occupazione tra laureati e non laureati. Si tratta di questioni centrali per la valutazione delle politiche legate al diritto allo studio e per l'analisi delle diseguaglianze economiche e sociali a livello territoriale.

Risultati preliminari

Il focus dell'analisi è la mobilità tra università e lavoro tra macro-aree. Tra i laureati magistrali, il 30% presenta almeno uno spostamento post-laurea rispetto alla macro-area dell'ateneo, mentre il 70% rimane stabilmente nella macro-area dell'ateneo. La mobilità è più contenuta tra i laureati triennali: il 23% sperimenta almeno uno spostamento, mentre il 77% rimane nell'area dell'ateneo. Il differenziale tra laureati magistrali e laureati triennali suggerisce che livelli di istruzione più elevati siano associati a una maggiore mobilità territoriale. La laurea magistrale sembra dunque ampliare l'orizzonte geografico del mercato del lavoro di riferimento.

Per analizzare più in dettaglio i singoli spostamenti, si mostrano di seguito le matrici di transizione dalla macro-area di conseguimento del titolo di studio universitario per i laureati (figura 2) o dell'ultimo ateneo frequentato per coloro che hanno abbandonato gli studi (figura 3) alla macro-area in cui ha sede l'azienda. Le unità statistiche sono qui costituite dai rapporti di lavoro;

Figura 2. Quote percentuali dei rapporti di lavoro per macro-area di partenza (sede dell'ateneo) e di arrivo (luogo di lavoro), per tipologia di laurea conseguita e livello di qualificazione del lavoro (classificazione ISCO)



Nota: Panel superiore: laurea magistrale, panel inferiore: laurea di primo livello. Coorti di immatricolazione 2009/10-2016/17. Rapporti di lavoro osservati negli archivi WHIP/COB fino al 2019.
Fonte: elaborazione degli Autori

pertanto, ogni individuo può essere coinvolto in più transizioni[7]. Le matrici sono articolate per livello di qualifica dell'occupazione, sulla base del primo digit del codice ISCO: occupazioni altamente qualificate (codici 1-3), mediamente qualificate (codici 4-6) e non qualificate (codici 7-9).

L'analisi degli spostamenti rivela che la propensione alla mobilità non è neutra rispetto alla qualità dell'inserimento occupazionale: gli spostamenti sono più frequenti nei segmenti professionali medio-alti, coerentemente con l'ipotesi che la mobilità sia parte di strategie di investimento in capitale umano e accesso a mercati del lavoro più dinamici.

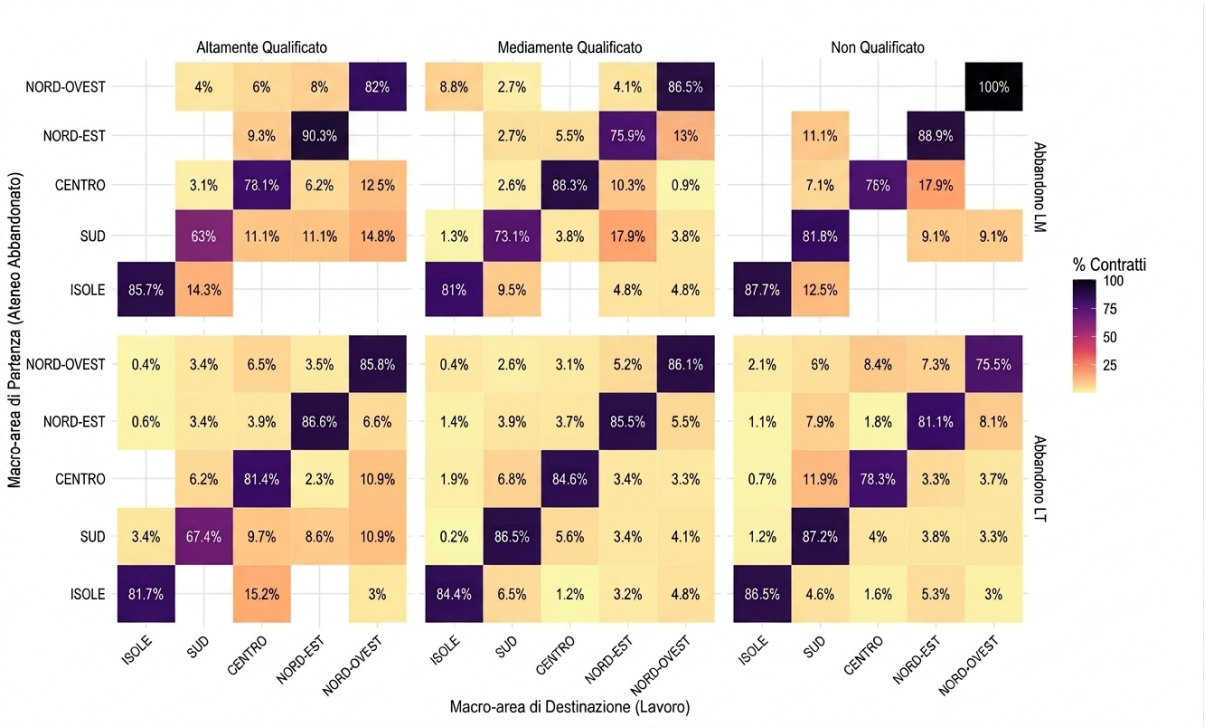
Emergono inoltre forti differenze tra le diverse aree geografiche, in particolare per i laureati. Mentre le regioni del Centro e del Nord mostrano un'elevata capacità di trattenere i propri laureati, con tassi di permanenza sempre largamente superiori al 70%, nel Sud e nelle Isole la dinamica è differente. Nello specifico, tra i laureati magistrali del Sud, oltre un contratto su due tra quelli ad alta qualifica (il 53,2%) viene stipulato in un'area diversa da quella di studio, con una marcata polarizzazione verso il Nord-Ovest.

In generale, il Centro-Nord e in particolare il Nord-Ovest si configurano come i principali poli di attrazione, anche per le professioni non qualificate. È il caso, ad esempio, dei laureati magistrali delle Isole: dell'8,9% che svolge un lavoro non qualificato (cfr. tabella A3 in Appendice), il 60% lavora in regioni del Nord-Ovest. Invece, i tassi di permanenza nella macroarea di origine relativi al lavoro mediamente qualificato superano ovunque il 75%, suggerendo una buona capacità di assorbimento di questo tipo di professioni da parte dei mercati locali.

Di particolare interesse è il fenomeno della mobilità di ritorno, con cui si identificano gli individui che – al termine della loro esperienza universitaria fuori sede – tornano a lavorare nella area di origine. Nel nostro campione, il 41,4% dei laureati triennali che avevano effettuato uno spostamento di area rientra nella macro-area di provenienza in almeno

7 Per le analisi sul lavoro dei laureati sono stati considerati i soli contratti stipulati almeno nello stesso anno di ottenimento della laurea o anni successivi, in modo da valutare l'effetto del conseguimento del titolo.

Figura 3. Quote percentuali dei rapporti di lavoro per macro-area di partenza (sede dell'ateneo) e di arrivo (luogo di lavoro) e livello di qualificazione del lavoro (classificazione ISCO), per gli studenti che hanno abbandonato il percorso universitario



Nota: Panel superiore: laurea magistrale, panel inferiore: laurea di primo livello. Coorti di immatricolazione 2009/10-2016/17. Rapporti di lavoro osservati negli archivi WHIP/COB fino al 2019.
Fonte: elaborazione degli Autori

una fase della carriera lavorativa. Tra gli studenti originari del Sud e delle Isole – che sperimentano la mobilità di studio di gran lunga più elevata – il tasso di ritorno è molto più basso rispetto a quello osservato nelle aree del Centro e del Nord: pari rispettivamente al 43,4% e 47,3% per i laureati triennali e al 26% e 29,3% per i laureati magistrali. Nelle aree del Centro e del Nord, invece, i tassi di ritorno si attestano intorno al 60-80% per i laureati triennali e tra il 43% e il 72% per i laureati magistrali (tabella A4 in Appendice).

Il grafico relativo agli abbandoni (figura 3) mostra un pattern strutturalmente diverso: chi abbandona il percorso universitario tende a restare nella macro-area dell'ateneo in misura maggiore rispetto a chi lo conclude. Per gli abbandoni di percorsi di laurea magistrale, i tassi di permanenza superano quasi ovunque l'80-85%, evidenziando una ridotta propensione alla mobilità tra coloro che non conseguono il titolo. I casi di mobilità risultano, anche in questo caso, polarizzati verso il Nord: in particolare, si osservano spostamenti dal Sud verso il Centro e il Nord-Ovest per lavori altamente qualificati e dal Sud verso il Nord-Est per quelli mediamente qualificati.

Nel caso degli abbandoni triennali, i tassi di permanenza sono più elevati e omogenei tra livelli di qualificazione. Fa eccezione, anche in questo caso, il Sud, dove la capacità di trattenere profili qualificati è sensibilmente minore. Va inoltre notato che i flussi verso professioni ad alta qualifica rappresentano una componente minoritaria del fenomeno: nel caso dei drop-out triennali, meno del 5% degli individui che hanno studiato al Sud riesce a inserirsi in segmenti lavorativi qualificati. Diversamente, al Nord oltre al 10% dei contratti è stipulato per lavori altamente qualificati, anche in assenza di una laurea (cfr. tabella A3 in Appendice).

In sintesi, il completamento del percorso universitario sembra modificare la natura della mobilità territoriale: nei laureati essa assume carattere selettivo e orientato verso segmenti qualificati del mercato del lavoro; tra coloro che hanno abbandonato il percorso di studi tende invece a configurarsi come mobilità meno strutturata e meno associata a occupazioni di elevata qualità, specialmente per chi non ha conseguito la laurea

triennale. L'interruzione degli studi, dunque, agisce da freno alla mobilità geografica.

Conclusioni

La sperimentazione presentata in questo contributo mostra come dati amministrativi prodotti da enti diversi possano essere integrati, anche in assenza di identificativi diretti, attraverso una procedura di *linkage* costruita nel rispetto della tutela della riservatezza. Attraverso un disegno a doppio cieco e l'uso di un set minimo di informazioni (sesso, data e comune di nascita), combinati con regole di esclusione temporali e biografiche basate sulla coerenza tra carriere accademiche e lavorative, è stato costruito un dataset di circa 17.000 match 1:1 ad alta affidabilità tra percorsi di studio e rapporti di lavoro. Questo sottoinsieme costituisce una base informativa utile per esplorare diversi temi di interesse, tra i quali la mobilità interna e il confronto tra percorsi di laureati e non laureati, con una profondità temporale e territoriale difficilmente raggiungibile con le rilevazioni attualmente disponibili. Al netto della forte dipendenza dalla dimensione del comune di nascita, l'appartenenza al gruppo 1:1 non è selettiva rispetto alle principali caratteristiche individuali e dei percorsi universitari considerati.

La struttura integrata del dataset consente infatti di sviluppare una lettura più granulare e dinamica delle traiettorie studio-lavoro, valorizzando appieno la dimensione longitudinale delle informazioni disponibili. In particolare, l'archivio ANS permette di seguire l'intero percorso universitario degli individui, dall'immatricolazione al conseguimento del titolo o all'abbandono, osservando esami sostenuti, crediti acquisiti e tempistiche di avanzamento. L'integrazione con i dati sulle carriere lavorative consente di collegare tali traiettorie formative agli esiti occupazionali successivi, aprendo la possibilità di analizzare non solo la qualità dell'inserimento lavorativo, ma anche la relazione tra progressione accademica, interruzione degli studi e dinamica delle carriere nel medio periodo.

In questa prospettiva, il lavoro presentato apre alcune linee di sviluppo, principalmente orientate a migliorare la qualità del *linkage* e l'utilizzabilità del dataset risultante.

Un primo ambito riguarda la validazione del dataset dei match 1:1. Il confronto con *linkage* deterministici realizzati su base locale, laddove disponibili, consentirebbe di valutare in modo più diretto la qualità dei match 1:1 e di definire criteri più precisi per la loro inclusione nelle analisi empiriche.

Un secondo ambito riguarda l'estensione dell'analisi oltre il sottoinsieme dei match univoci. Il dataset 1:1 rappresenta infatti solo una parte dell'informazione disponibile. Una direzione promettente consiste nello sviluppo di strategie di assegnazione probabilistica e di imputazione multipla, basate su informazioni ausiliarie provenienti da altre fonti statistiche, per utilizzare anche i match non univoci (1:molti, molti:1 e molti:molti), incorporando così esplicitamente l'incertezza nelle analisi empiriche.

Infine, un aspetto cruciale riguarda la costruzione di infrastrutture dati stabili che rendano possibile, su base continuativa, l'integrazione tra archivi di istruzione e lavoro a fini di ricerca. La sperimentazione qui presentata dimostra la fattibilità tecnica di questa integrazione in un contesto privacy-preserving, ma evidenzia al tempo stesso la necessità di sviluppare soluzioni istituzionali e organizzative che ne consentano l'utilizzo sistematico su scala nazionale.

Nel complesso, il lavoro suggerisce che il valore analitico dei dati amministrativi non risiede solo nella loro disponibilità, ma nella possibilità di integrarli in modo metodologicamente solido, esplicito nei limiti e trasparente nelle assunzioni. In questa prospettiva, il *linkage* tra ANS e WHIP rappresenta non solo uno strumento per l'analisi empirica, ma anche un banco di prova per lo sviluppo di nuove infrastrutture di ricerca basate su dati amministrativi.

Appendice

Tabella A1. Tasso di risposta nell'indagine AlmaLaurea 2022 ai laureati

	N. laureati	% di risposta sul totale dei laureati	% di risposta sul totale dei laureati contattabili
A un anno			
Primo livello	167.467	68,7	72,6
Magistrali biennali	94.673	67,1	73,9
Magistrali a ciclo unico	34.257	69,2	73,9
A tre anni			
Primo livello	72.945	15,2	15,2
Magistrali biennali	86.117	53,7	76,8
Magistrali a ciclo unico	35.969	68,9	75,8
A cinque anni			
Primo livello	61.490	10,2	10,2
Magistrali biennali	80.369	55,3	64,5
Magistrali a ciclo unico	36.082	54,1	63,6
Sc. Formazione (pre-riforma)	409	54,8	60,4

Fonte: AlmaLaurea, *XXV Indagine Condizione occupazionale dei laureati*

Tabella A2. Probabilità di appartenenza al gruppo 1:1 rispetto a match non univoci. Risultati da modello logit

	(1)			(2)		
Variabile	**AME**	**SE**	**p-value**	**AME**	**SE**	**p-value**
Genere maschile	0,0291	0,0035	0,0000	0,0284	0,0034	0,0000
Coorte imm. (rif 2009)						
2010	-0,0045	0,0061	0,4569	-0,0046	0,0061	0,4522
2012	0,0024	0,0060	0,6903	0,0025	0,0061	0,6846
2013	-0,0257	0,0060	0,0000	-0,0250	0,0061	0,0000
2014	-0,0242	0,0061	0,0001	-0,0224	0,0064	0,0005
2015	-0,0298	0,0061	0,0000	-0,0275	0,0064	0,0000
2016	-0,0159	0,0064	0,0132	-0,0137	0,0069	0,0453
Tipo diploma (rif. Altri licei)						
Classico	-0,0106	0,0063	0,0920	-0,0114	0,0063	0,0676
Scientifico	-0,0113	0,0049	0,0217	-0,0119	0,0049	0,0141
Professionale	0,0070	0,0077	0,3634	0,0071	0,0077	0,3597
Tecnico	-0,0023	0,0053	0,6612	-0,0030	0,0052	0,5690
Estero/Altro	-0,0246	0,0141	0,0803	-0,0256	0,0141	0,0682
Dim. Comune (rif.<10 mila)						
10-25 mila	-0,0875	0,0109	0,0000	-0,0873	0,0109	0,0000
25-50 mila	-0,1883	0,0103	0,0000	-0,1885	0,0103	0,0000
50-100 mila	-0,3416	0,0103	0,0000	-0,3418	0,0103	0,0000
100-250 mila	-0,5093	0,0101	0,0000	-0,5095	0,0101	0,0000
250-500 mila	-0,6515	0,0098	0,0000	-0,6517	0,0098	0,0000
500-3 milioni	-0,7164	0,0090	0,0000	-0,7168	0,0090	0,0000

segue

segue tabella A2

Area disciplinare (rif. Sanitaria)						
Economico-giuridica	0,0071	0,0074	0,3337			
Formazione	0,0321	0,0115	0,0051			
Ingegneria-Architettura	0,0099	0,0076	0,1921			
Politico-sociale	0,0067	0,0090	0,4562			
Scienze-tecnologia	0,0271	0,0078	0,0005			
Umanistico-linguistica	0,0156	0,0075	0,0384			
Non laureati	0,0107	0,0059	0,0729			
Stato percorso (rif. Laurea Triennale in corso)						
Laurea Triennale				-0,0031	0,0052	0,5433
Laurea Magistrale				0,0093	0,0060	0,1210
Laurea Magistrale in corso				0,0051	0,0057	0,3728
Abbandono Laurea Triennale				0,0035	0,0050	0,4827
Abbandono Laurea Magistrale				0,0113	0,0143	0,4307
Passaggio a Laurea a Ciclo Unico				-0,0232	0,0187	0,2145

Fonte: elaborazione degli Autori

Tabella A3. Qualifica dell'occupazione per area geografica dell'Ateneo e percorso universitario

	Altamente qualificato	Mediamente qualificato	Non qualificato
		Laureati magistrali	
Nord-Ovest	63,2	30,9	5,9
Nord-Est	56,3	34,5	9,2
Centro	56,6	34,6	8,8
Sud	55,1	35,5	9,4
Isole	56,3	34,8	8,9
		Laureati triennali	
Nord-Ovest	36,4	50,2	13,4
Nord-Est	26,3	57,0	16,7
Centro	26,5	61,0	12,5
Sud	22,4	59,0	18,6
Isole	23,6	55,1	21,3
		Abbandoni magistrali	
Nord-Ovest	38,2	56,5	5,3
Nord-Est	33,0	57,4	9,6
Centro	18,1	66,2	15,7
Sud	23,3	67,3	9,4
Isole	19,5	58,4	22,1
		Abbandoni triennali	
Nord-Ovest	12,3	69,9	17,8
Nord-Est	9,3	67,6	23,1
Centro	6,5	72,9	20,6
Sud	5,5	68,2	26,3
Isole	3,5	64,0	32,5

Fonte: elaborazione degli Autori

Tabella A4. Quota di studenti fuori sede e tasso di ritorno al termine degli studi per area di origine

Area di origine	% Fuori sede	Tasso di ritorno
Laureati triennali		
Nord-Ovest	13,8	79,4
Nord-Est	9,8	58,3
Centro	13,8	67,6
Sud	40,2	43,4
Isole	22,4	47,3
Laureati magistrali		
Nord-Ovest	8,0	72,5
Nord-Est	8,7	47,7
Centro	16,1	43,2
Sud	48,9	26,0
Isole	18,3	29,3
Abbandoni triennali		
Nord-Ovest	11,0	63,9
Nord-Est	6,4	90,5
Centro	11,3	67,6
Sud	54,6	61,8
Isole	16,8	62,0
Abbandoni magistrali		
Nord-Ovest	9,1	33,3
Nord-Est	30,3	70,0
Centro	9,1	66,7
Sud	39,4	25,0
Isole	12,1	53,8

Fonte: elaborazione degli Autori

Bibliografia

Ardito C., Cottini E., Moroni G., Serrano-Alarcon M. (2026), *Ageing on the Job: How Raising Retirement Age Affects Work Accidents*, WorkInps paper n.110, gennaio

Attanasio M., Ragozini G., Porcu, M., Giambalvo O. (2020), *Verso nord: Le nuove e vecchie rotte delle migrazioni universitarie*, Milano, Franco Angeli

Berton F., Richiardi M., Sacchi, S. (2012), *The political economy of work security and flexibility: Italy in comparative perspective*. Policy Press, Bristol

D'Agostino A., Ghellini G., Longobardi S. (2019), Out-migration of university enrolment: the mobility behaviour of Italian students, *International Journal of Manpower*, 40, n.1, pp.56-72

De Angelis I., Mariani V., Torrini R. (2017), New evidence on interregional mobility of students in tertiary education: The case of Italy, *Politica economica*, 33, n.1, pp.73-96

Devicienti F., Grinza E., Manello A., Vannoni D. (2025), Employer cooperation, productivity and wages: new evidence from inter-firm formal network agreements, *Economica*, 9, n.365, 1-41

Dotti N. F., Fratesi U., Lenzi C., Percoco M. (2013), Local labour markets and the interregional mobility of Italian university students, *Spatial Economic Analysis*, 8, n.4, pp.443-468

Galizzi M., Leombruni R., Pacelli L. (2023), Severe work disabilities and long-lasting losses, *Labour Economics*, 85, Artcle n.102432, <https://doi.org/10.1016/j.labeco.2023.102432>

Leombruni R. (2024), Interviewing administrative records. A conceptual map for the use of big data for economic research, *Italian Journal of Applied Statistics*, 36, n.3, pp.295-326

Leombruni R., Ardito C. (2026), The long and winding road. The access to administrative data for research in Italy, in Berton F., Paruolo P. (a cura di), *Data-Driven Learning in the EU. From Administrative Sources to Public Value*, Springer Cham, Joint Research Centre

Priulla A., Miaci E., Panichella N. (2025), Social selection in student mobility: The interplay of origin, achievement and track in the Italian North-South divide, *Research in Social Stratification and Mobility*, 100, n.101105, pp.1-10

Tosi F., Impicciatore R., Rettaroli R. (2025), Italian graduates' geographical mobility patterns: selectivity and regional inequalities, *Higher Education*, n.91, pp.869-893

Usala C., Porcu M., Sulis I. (2023), The high school effect on students' mobility choices, *Statistical Methods & Applications*, 32, n.4, pp.1259-1293

Andrea Priulla
andrea.priulla@unikore.it
Ricercatore in Statistica presso il Dipartimento di Scienze economiche e giuridiche dell'Università degli Studi di Enna "Kore". La sua attività di ricerca si concentra sulle disuguaglianze socioeconomiche, di genere, etniche e territoriali nell'istruzione universitaria italiana. Tra le pubblicazioni recenti si segnalano: con Albano A., D'Angelo N., Attanasio M., Predicting university enrollment choices in Italy: a machine learning analysis of high school background and gender differences, *Higher Education*, 2025 e con Miaci E., Panichella N., Social Selection in Student Mobility, The Interplay of Origin, Achievement and Track in the Italian North-South Divide, *Research in Social Stratification and Mobility*, 2025.

Dalit Contini
dalit.contini@unito.it
Professoressa ordinaria di Statistica sociale presso il Dipartimento di Economia e statistica "Cognetti de Martiis" dell'Università di Torino. Si occupa di disuguaglianze educative, istruzione terziaria, valutazione di politiche educative. Tra le pubblicazioni recenti si segnalano: con Di Tommaso M.L., Muratori C., Piazzalunga D., Schiavon L., A lost generation? Impact of Covid-19 on high school students' achievements, *Journal of Economic Inequality*, 2025 e con Salza G., Children left behind. New evidence on the (adverse) impact of grade retention on educational careers, *European Sociological Review*, 2024.

Sonia Della Monica
sonia.dellamonica@unito.it
Assegnista di ricerca presso il Dipartimento di Economia e statistica "Cognetti de Martiis" dell'Università di Torino. Fra le pubblicazioni recenti si segnalano: con Leombruni R., Infortuni lavorando, ma senza tutele. Nuove evidenze su sotto-copertura e sotto-notifica degli infortuni sul lavoro in Italia, *Economia & lavoro*, 2026 e con Storti L., Going up: Romanian Immigrants in an Alpine Region, *Mondi Migranti*, 2026.

Noemi Ferrara
noemi.ferrara@edu.unito.it
Borsista di ricerca presso il Dipartimento di Economia e statistica "Cognetti de Martiis" dell'Università di Torino e ricercatrice junior presso l'Istituto di Ricerche economiche e sociali del Piemonte (IRES).

Integrazione dei brevetti e delle imprese italiane

Sergio Salamone
ISTAT

Il lavoro propone un'analisi data-driven sull'uso di dati amministrativi per la tutela brevettuale integrati con informazioni sulle imprese. I dati dei brevetti offrono indicazioni sui processi innovativi attivati dalle imprese e la relazione con le loro caratteristiche strutturali ed economiche suggerisce possibili strategie sulla propensione a brevettare a parità di profili aziendali. Le informazioni integrate offrono un'opportunità di analisi micro-fondata per le policy legate all'innovazione, competitività, e transizione tecnologica.

Integration of patents and Italian firms

The paper presents a data-driven analysis of the use of administrative data for patent protection, integrated with firm-level information. Patent data provide insights into the innovation processes undertaken by firms, and the relationship with firms' structural and economic characteristics suggests potential strategies regarding the propensity to patent among companies with similar profiles. The integrated information offers an opportunity for micro-founded analysis to support policies related to innovation, competitiveness, and technological transition.

DOI: 10.53223/Sinappsi_2026-02-8

Citazione
Salamone S. (2026), Integrazione dei brevetti e delle imprese italiane, *Sinappsi*, XVI, pp.122-134

Parole chiave
Dati amministrativi
Imprese
Innovazioni tecnologiche

Keywords
Administrative data
Enterprises
Technological innovation

Introduzione

Negli ultimi anni i dati amministrativi hanno assunto un ruolo centrale nelle politiche pubbliche, poiché consentono di basare le decisioni su evidenze empiriche e di disporre di informazioni aggiornate, affidabili e a basso costo. L'importanza dei dati amministrativi per le politiche è testimoniata dalle crescenti attività in particolare della Ragioneria generale dello Stato, l'Ufficio parlamentare di Bilancio (UPB), la Corte dei conti, ma anche gli uffici del Senato e della Camera. La loro integrazione nei processi decisionali favorisce interventi più mirati ed efficaci (Azzone 2018). Le amministrazioni pubbliche producono, raccolgono e integrano grandi quantità di dati (sanità, imprese, ricerca, lavoro, istruzione, mobilità), caratterizzati da continuità di aggiornamento, standardizzazione, copertura dell'intera popolazione e natura longitudinale. Pur essendo raccolti per finalità amministrative, rappresentano una risorsa fondamentale per l'analisi e la valutazione delle politiche (Card *et al.* 2010).

Una questione rilevante riguarda la disponibilità di dati tempestivi per comprendere ex ante il contesto di riferimento per gli interventi di policy, o ex post per la valutazione con informazioni disegnate in maniera coerente con gli obiettivi dell'intervento. Tale disponibilità assieme a quella di micro-dati a livello di individuo o impresa consente non solo il monitoraggio degli interventi, ma anche la valutazione dei loro effetti, distinguendo gruppi target e gruppi di controllo (Crato e Paruolo 2019). Inoltre, l'integrazione tra dati amministrativi, survey e altre fonti informative accresce il potenziale conoscitivo e il supporto alle decisioni pubbliche (von Lucke e Frank 2025). Tali integrazioni si basano su metodologie consolidate sviluppate nell'ambito della statistica ufficiale (Scanu 2013) e della gestione di basi dati amministrative (Jabine e Scheuren 1986).

Uno strumento metodologico chiave è il *record*

linkage, che permette di collegare informazioni provenienti da fonti diverse in assenza di identificativi univoci, utilizzando variabili di matching (Shlomo 2019). L'uso di dati integrati è cresciuto notevolmente negli ultimi anni (Connelly *et al.* 2016), anche grazie all'impiego di open data, web scraping e text mining (Barcaroli *et al.* 2015). Parallelamente, i progressi tecnologici e metodologici, inclusi i metodi controfattuali per la valutazione d'impatto, hanno ampliato le possibilità di utilizzo dei dati per le policy (Athey *et al.* 2020).

In questo contesto, i dati brevettuali rappresentano una fonte amministrativa particolarmente rilevante per lo studio dell'innovazione. Oggi sono accessibili e analizzabili attraverso strumenti di text mining, machine learning e intelligenza artificiale. Integrati con le indagini statistiche su ricerca, sviluppo e innovazione (ad esempio le indagini Istat su R&S e innovazione o l'indagine sugli asset intangibili di Inapp), consentono una comprensione più completa dei processi innovativi, della proprietà intellettuale e la definizione di statistiche integrate sull'economia della conoscenza (OECD 2015; OECD e Eurostat 2018).

Il presente lavoro integra le informazioni sulle domande di brevetto con le caratteristiche economiche e strutturali delle imprese con almeno un deposito tra il 2015 e il 2023. L'obiettivo è la definizione di un sistema informativo come strumento utile per le politiche di ricerca, innovazione e trasferimento tecnologico, che consente di identificare specifici gruppi di imprese innovatrici. La principale implicazione per le policy della disponibilità di informazioni a livello di micro-dato sulle imprese con brevetti riguarda l'individuazione di specifici target di aziende identificabili e selezionabili sui quali orientare e valutare gli interventi.

L'integrazione permette infatti di analizzare il diverso contributo di PMI e grandi imprese all'attività brevettuale, ad esempio la relazione tra concentrazione brevettuale e concentrazione di mercato che consegna l'innovazione nelle mani di pochi grandi player. Le relazioni tra innovazione, produttività e occupazione, nonché l'efficacia delle collaborazioni tra imprese e organismi di ricerca per le politiche sul trasferimento tecnologico.

Poiché i brevetti fungono da segnale anticipatore di trend tecnologici emergenti, le caratteristiche strutturali ed economiche delle imprese brevettanti e i settori di attività in cui operano danno suggerimenti su nicchie strategiche o potenziali di innovazione su cui poter intervenire prima che l'innovazione entri nel mercato, come per la recente diffusione dei servizi di intelligenza artificiale e dei comparti ad essa associata nella produzione di componenti software, micro e nanoelettronica.

Il lavoro è strutturato partendo dall'inquadramento dei brevetti come misure per l'innovazione. Vengono descritte le fonti dei dati utilizzate, le metodologie e il processo di integrazione tra le domande di brevetto e le imprese (applicant) che hanno depositato i brevetti nel periodo dal 2015 al 2023. I risultati declinati sulla base delle caratteristiche strutturali delle imprese, per settore di attività economica e specializzazioni tecnologiche. Infine, vengono suggeriti collegamenti tra l'output e l'utilizzo per le policy legate alla transizione tecnologica, R&S e innovazione.

1. Contesto di riferimento e letteratura

In letteratura i brevetti sono considerati tra i principali indicatori dell'innovazione, in particolare delle attività inventive e creative delle imprese. Le domande di brevetto costituiscono una fonte oggettiva, omogenea e standardizzata, gestita dagli uffici nazionali ed europei per la tutela della proprietà intellettuale (Lotti e Nobile 2025). Il loro utilizzo come proxy dell'innovazione deriva dalla difficoltà di osservare direttamente la conoscenza prodotta e lo sforzo inventivo delle imprese: i brevetti rappresentano infatti una delle più importanti tracce amministrative dell'attività innovativa.

Nella letteratura empirica, i brevetti sono impiegati come misura dell'innovazione codificata e formalizzata, distinta dall'innovazione tacita composta da miglioramenti incrementali che possono accumularsi fino a generare invenzioni brevettabili (Boix-Domènech *et al.* 2025). Essi non misurano l'innovazione piena in senso schumpeteriano, ma l'attività inventiva legalmente appropriabile; da qui il problema della misurazione dell'innovazione, per cui la teoria è più sviluppata degli strumenti empirici disponibili (Schmookler 1966).

I primi studi hanno evidenziato una relazione tra brevetti e performance economiche, interpretando i brevetti come output intermedio del processo R&S → conoscenza → brevetti → produttività (Pakes e Griliches 1984). Da questa prospettiva nasce l'idea di utilizzare il conteggio dei brevetti come misura della capacità inventiva e innovativa (Griliches 1998).

La letteratura evidenzia tuttavia importanti limiti: i brevetti misurano l'invenzione e non necessariamente l'innovazione introdotta sul mercato (Pavitt 1984); molte innovazioni non vengono brevettate perché protette tramite segreto industriale; inoltre, il valore economico e tecnologico dei brevetti è molto eterogeneo, con pochi brevetti di grande impatto e molti di valore limitato (Trajtenberg 1990). Per questo motivo sono stati sviluppati indicatori e linee guida

per misurare la qualità dei brevetti e costruire statistiche comparabili (OECD 2009). Tali studi hanno favorito analisi sul contributo dei brevetti alle performance delle imprese (Squicciarini *et al.* 2013) e l'integrazione dei dati brevettuali con archivi aziendali (AIDA, ORBIS, AMADEUS).

Tra le applicazioni italiane, Lotti e Schivardi (2005) hanno collegato le domande di brevetto EPO alle imprese, evidenziando margini di miglioramento nella precisione degli abbinamenti, soprattutto per il mancato utilizzo delle informazioni territoriali. Successivamente, metodi basati su regole, dizionari e processi di standardizzazione specifici hanno migliorato l'integrazione dei dati (Thoma *et al.* 2010). Un'analisi successiva ha inoltre descritto la distribuzione degli abbinamenti impresa-brevetto per periodo, settore, dimensione aziendale, area geografica e dominio tecnologico (Lotti e Marin 2013).

Negli ultimi anni, la disponibilità e la standardizzazione delle informazioni brevettuali sono notevolmente migliorate (EPO 2025). Parallelamente, i progressi nelle tecniche di text mining e Natural language processing (NLP) hanno ampliato le possibilità di ricerca, analisi e utilizzo dei dati sui brevetti, rendendoli una fonte sempre più rilevante per lo studio dell'innovazione e per il supporto alle politiche pubbliche[1].

2. Fonti, dati e metodologie

European patent office (EPO)

Le informazioni pubblicate dall'Ufficio europeo dei brevetti (EPO) riguardano le domande e le concessioni di brevetto che hanno superato le verifiche necessarie per ottenere la tutela dell'invenzione. Un brevetto attribuisce un diritto esclusivo su un prodotto o processo innovativo che presenti utilità pratica, novità rispetto allo stato dell'arte e conformità ai requisiti normativi. La protezione dura fino a 20 anni, durante i quali il titolare può concedere licenze o trasferire il brevetto; alla scadenza, l'invenzione entra nel pubblico dominio. Il processo amministrativo per la brevettazione inizia con il deposito della domanda e prevede la pubblicazione di diversi documenti che ne descrivono l'avanzamento: i documenti A corrispondono alle domande di brevetto, mentre i documenti B alle concessioni. In questo report si considerano esclusivamente le domande di brevetto. La pubblicazione della domanda avviene generalmente dopo 18 mesi dalla data di deposito, mentre la concessione può richiedere diversi anni. Nel processo intervengono tre soggetti principali: inventore, titolare e richiedente (applicant), che possono non coincidere.

Questo lavoro si concentra sui richiedenti con sede in Italia che hanno depositato almeno un brevetto nel periodo 2015-2023; tali soggetti possono essere individui, imprese, università o istituzioni.

L'EPO pubblica obbligatoriamente sia le domande sia i brevetti concessi attraverso l'European publication server (EPS), fonte ufficiale dei documenti europei dal 2005, ed utilizzata per questo lavoro. Le informazioni utilizzate comprendono dati anagrafici e di localizzazione di inventori, richiedenti e titolari, date e numeri di priorità, deposito, pubblicazione e concessione, nonché titolo, abstract, classi IPC, citazioni e *claims* del brevetto.

Registri statistici sulle imprese

La fonte informativa sulle imprese è costituita dai registri statistici Istat, in particolare ASIA (Archivio statistico delle imprese attive) e ASIA Unità locali (UL). Quest'ultimo è stato utilizzato poiché consente di considerare anche le sedi operative diverse dalla sede legale, rilevanti nel caso in cui le attività di ricerca e sviluppo che generano brevetti siano svolte presso specifiche unità locali.

È stato costruito un panel di imprese ASIA UL 2015-2023, utilizzando per il *record linkage* le variabili anagrafiche disponibili in entrambe le fonti: ragione sociale, comune, indirizzo e CAP. Per gestire variazioni nel tempo delle informazioni aziendali (ad esempio cambi di indirizzo), sono state considerate tutte le versioni annuali delle unità locali, migliorando la probabilità di identificazione e lo score di abbinamento.

Integrazione dei brevetti con i registri statistici tramite record linkage

L'integrazione tra i dati brevettuali EPO e i micro-dati delle imprese del registro statistico ASIA richiede una procedura di collegamento, poiché le due fonti non dispongono di un identificativo univoco comune (ad esempio il codice fiscale) che consenta un abbinamento diretto.

La metodologia adottata è il *record linkage*, finalizzato a individuare coppie di record appartenenti alla stessa unità reale attraverso l'utilizzo congiunto di variabili identificative. La procedura algoritmica mira quindi a riconoscere le imprese presenti in entrambe le basi dati e si articola nelle fasi di Pre-linkage, Data linkage e Post-linkage, applicate in modo sequenziale e iterativo con verifiche visive e manuali dei risultati. Per l'implementazione è stato utilizzato il *record linkage* toolkit di Python, scelto per flessibilità e disponibilità di diversi strumenti di confronto.

1 Si veda https://link.epo.org/web/searching-for-patents-brochure_en.pdf.

Le variabili utilizzate per l'abbinamento sono state selezionate tra quelle disponibili in entrambe le fonti: ragione sociale, Comune, CAP e indirizzo. Queste informazioni, opportunamente standardizzate nella fase di pre-processing, assumono congiuntamente il ruolo di identificativo dell'impresa. La ragione sociale è stata utilizzata come variabile principale di blocking per ridurre lo spazio di ricerca degli abbinamenti.

Per la localizzazione, CAP e Comune sono stati utilizzati in modo complementare, considerando eventuali valori mancanti o criticità legate ai CAP generici delle aree urbane con più codici postali.

Pre-processing, cleaning e standardizzazione

La qualità del *record linkage* dipende dalla qualità dei dati di partenza. Per questo motivo è stata dedicata particolare attenzione alla fase di pre-processing, soprattutto alla standardizzazione della ragione sociale del richiedente (applicant), variabile principale per l'abbinamento. La pulizia dei dati ha consentito di ridurre ambiguità, come la presenza di numerose imprese con la stessa denominazione nei registri Istat.

La fase preliminare ha riguardato l'estrazione della forma giuridica dalla ragione sociale e la standardizzazione delle variabili di matching (ragione sociale, comune, indirizzo e CAP) in entrambe le fonti, tramite rimozione di elementi non informativi (token, spazi, parentesi e caratteri indesiderati).

Per la ragione sociale sono state applicate regole specifiche per uniformare abbreviazioni e forme giuridiche (es. Spa), eliminando anche parentesi e informazioni accessorie presenti nei nominativi EPO e ASIA. Successivi controlli visivi hanno portato a ulteriori armonizzazioni delle denominazioni societarie.

Nel secondo step della procedura Neighboroud per la ragione sociale, le analisi visive hanno suggerito la necessità di effettuare un ulteriore standardizzazione trasformando le forme giuridiche esplicitate (ad esempio per le Spa la dicitura in forma abbreviata Spa).

Per il Comune, i dati EPO sono stati puliti eliminando riferimenti alla provincia, anche presenti tra parentesi, e uniformando le denominazioni tradotte dall'inglese all'italiano (es. Milan → Milano, Florence → Firenze, Rome → Roma).

Per l'indirizzo, utilizzato soprattutto per distinguere imprese con stessa ragione sociale nella medesima area geografica, sono state rimosse informazioni non rilevanti come frazioni, località o indicazioni aggiuntive.

Selezione delle imprese con brevetto da EPO

Preliminarmente alla riduzione dello spazio di abbinamento vero e proprio sono state ridimensionate entrambe le fonti EPO e ASIA per migliorare la capacità computazionale ed eliminare informazioni che aumentano la distorsione nell'abbinamento dei record. La fase di integrazione è preceduta dalla riduzione delle unità che brevettano da EPO attraverso la distinzione tra le tipologie di applicant, selezionando la popolazione target rappresentata dalle imprese ed escludendo le altre principali tipologie (persone fisiche, università, istituti, fondazioni, enti e centri di ricerca, consorzi, strutture ospedaliere).

La distinzione tra imprese e le altre tipologie di richiedenti è fondamentale per ridurre lo spazio di abbinamento nel *record linkage* e diminuire i casi di falsi abbinamenti.

Le domande di brevetto acquisite sono 35.470 e includono tutte le tipologie di applicant, quindi oltre le imprese anche istituzioni e individui. Gli applicant di EPO sono stati deduplicati, raggruppando i richiedenti con più brevetti, per ottenere un singolo richiedente. Nel periodo tra il 2015 e il 2023 risultano quindi 16.492 richiedenti (tabella 1). Le imprese rappresentano la porzione maggiore di richiedenti un brevetto, più del 70%, e vengono preliminarmente selezionate nel processo di integrazione dei dati per favorire il match con i registri statistici.

Il primo step per l'identificazione delle imprese con brevetti prevede l'analisi testuale del nome

Tabella 1. Numerosità degli applicant e dei brevetti per tipologia di richiedente (2015-2023)

Tipologia di Applicant	Applicant	Applicant (%)	Domande di brevetto (%)
Imprese	11.892	72,1%	88,7%
Persone fisiche	4.136	25,1%	6,2%
Università	218	1,3%	3,1%
Istituti, fondazioni, enti e centri di ricerca, consorzi	215	1,3%	1,6%
Strutture ospedaliere	31	0,2%	0,1%
Totale	16.492	100%	100%

Fonte: elaborazione dell'Autore su dati EPO

richiedente per selezionare le imprese attraverso la forma giuridica. Le imprese distinte per forma giuridica formano la popolazione di partenza per l'integrazione con i registri statistici.

L'individuazione delle imprese prevede ulteriori attività per casi particolari, distinguendo ad esempio tra imprese ed individui nei casi di imprese individuali (con ragione sociale del tipo NOME&COGNOME&DITTAINDIVIDUALE) rispetto alla struttura standard dei richiedenti come persone fisiche (COGNOME, NOME).

Selezione delle imprese dei registri statistici da abbinare con EPO

L'archivio delle Unità locali fornisce, oltre alle unità legali delle imprese, anche le unità operative localizzate sul territorio con finalità produttive o di supporto alla produzione, quindi la mole di informazioni utilizzate nel matching con i brevetti aumenta, in particolare quando un'impresa è attiva in più anni. Poiché la dimensione di questa popolazione influenza lo spazio di abbinamento definito per il *record linkage* con effetto moltiplicativo, il panel di imprese 2015-2023 (circa 13 milioni di record) è stato costruito selezionando solo le imprese con variazioni nelle variabili di matching, escludendo duplicati con corrispondenza esatta nelle variabili ragione sociale, Comune, CAP, indirizzo. Le variazioni presenti tra gli anni in almeno una di queste variabili sono utilizzate per ricercare la corrispondenza migliore con le variabili di matching in EPO.

Standardizzazione e codifica delle variabili di matching

Nella classificazione tramite *record linkage* delle imprese con domande di brevetti, assieme alla pulizia delle variabili di matching, assume rilevanza fondamentale la loro trasformazione attraverso metodi di codifica fonetica che permettono di superare differenze o errori nelle stringhe delle variabili di abbinamento. La codifica attraverso l'uso di codici fonetici compensa le differenze tra le stringhe nelle due fonti come nei casi delle forme giuridiche, ad esempio SPA rispetto a Società per azioni.

Per standardizzare le variabili di matching sono stati utilizzati i metodi match_rating, nysiis e soundex. I risultati migliore dono stati ottenuti con il metodo match_rating. Questi metodi sono stati applicati a tutte le variabili di abbinamento nelle due fonti e testati nelle fasi ricorsive di match per valutare quelli migliori.

Riduzione dello spazio di abbinamento, indexing e comparazione

La strategia adottata mira a ridurre lo spazio di ricerca tra le due fonti informative, limitando i confronti alle coppie di record con maggiore probabilità di riferirsi alla stessa impresa. A tal fine è stata applicata una procedura di indicizzazione per individuare le imprese richiedenti brevetti, presenti in entrambi i dataset. Dopo la creazione di un archivio integrato, sono stati applicati in modo ricorsivo diversi modelli di integrazione, identificando progressivamente le unità più simili non già abbinate nei passaggi precedenti. L'indicizzazione è stata realizzata definendo una variabile di *blocking* e le variabili di confronto. Come variabile di blocco è stata utilizzata la ragione sociale, la più completa e discriminante nelle due fonti. Tuttavia, poiché possono esistere imprese con denominazioni identiche o molto simili, sono state utilizzate anche le variabili di localizzazione per individuare l'abbinamento corretto.

Per la generazione delle coppie candidate è stato impiegato il toolkit *record linkage* di Python, utilizzando prima il Block Index per la selezione dei record con valori identici nella variabile di blocco, poi il Sorted Neighbourhood Index che confronta record ordinati per blocchi di dimensione prefissata (Christen 2012). L'abbinamento è stato effettuato con una procedura sequenziale uno-a-uno, basato su approccio deterministico prima più restrittivo sulla ragione sociale e successivamente sui record residui. Gli abbinamenti sono stati valutati tramite uno score di similarità e classificati come match o non match.

Poiché il numero di imprese brevettanti è relativamente contenuto, si è scelto di privilegiare la completezza degli abbinamenti rispetto all'efficienza computazionale, aumentando il numero di confronti per ridurre il rischio di perdere corrispondenze effettive.

A partire dalle coppie di imprese selezionate nella fase di indicizzazione, per valutare la corrispondenza dei record sono state applicate tecniche di comparazione delle stringhe e tuning dei parametri di abbinamento. In ogni step del processo iterativo sui record selezionati come *'candidate links'*, sono state usate metriche di similarità delle stringhe per gestire le differenze nelle variabili di matching.

Un comparatore di stringhe fornisce valori compresi tra 0, completa discordanza e 1, corrispondenza perfetta (Jaro 1989). Per confrontare le variabili di abbinamento sono state testate le misure Jaro, Jaro-Winkler e Levenshtein. Le sperimentazioni e controlli visivi suggeriscono il metodo Jaro-Winkler, poiché fornisce i risultati migliori, grazie alla capacità di attribuire maggiore peso alle stringhe con gli stessi caratteri iniziali e di produrre un punteggio di similarità normalizzato (Winkler 1990).

Classificazione con metodo basato su soglie (threshold-based method)

La classificazione deterministica è stata realizzata definendo soglie di similarità per distinguere i record *in match* e *non match*. Poiché questo approccio, pur semplice ed efficace, può generare errori sistematici, è stata affiancata un'estesa attività di revisione manuale per garantire la qualità dell'integrazione. La metrica utilizzata è uno score complessivo ottenuto dalla somma dei punteggi delle quattro variabili di matching (da 1 a 4), con un peso predominante attribuito alla ragione sociale. Le variabili di localizzazione (Comune, indirizzo e CAP) hanno contribuito a distinguere casi ambigui. Le soglie sono state definite attraverso analisi visive e verificate anche con il metodo ECM (Expectation/Conditional maximization).

Nel primo passaggio di abbinamento (Block Index), basato sulla coincidenza della ragione sociale e su una soglia di similarità Jaro-Winkler pari a 0,8, sono stati individuati 20.460 potenziali collegamenti. Di questi, il 65% dei record non duplicati è stato classificato come match, corrispondente a 5.542 imprese brevettanti, con score massimo pari a 4. La qualità degli abbinamenti è stata verificata tramite controlli manuali su campioni casuali (blocchi di cinquanta abbinamenti ordinati in base allo score complessivo). Gli abbinamenti con score complessivo uguale a 4 risultano dall'analisi visiva correttamente associati. Nello step successivo con l'applicazione del Sorted Neighbourhood Index tramite analisi visiva è stata identificato un punteggio maggiore di 3,7 complessivo di accettazione.

La revisione manuale ha rappresentato una fase centrale del processo. Ha consentito di ridurre i falsi positivi, affinare i parametri, migliorare la standardizzazione delle variabili e aumentare il numero di abbinamenti validi. La validazione ha combinato criteri di similarità testuale e coerenza geografica, adottando un approccio prudente volto a minimizzare errori di attribuzione senza rinunciare a massimizzare le corrispondenze. In particolare, la validazione visiva ha previsto il controllo simultaneo delle quattro variabili su ogni abbinamento ordinato per lo score complessivo. Sono stati confrontati i nomi nei due dataset e nella versione dopo la procedura di cleaning, affiancati a quelli standardizzati con nome armonizzato tramite algoritmi diversi (match rating, nysiis, soundex).

La validazione manuale ha seguito una linea conservativa per minimizzare l'associazione di brevetti ad imprese che non ne hanno depositato nel periodo considerato, ma nello stesso tempo questo lavoro mira a massimizzare il numero di corrispondenze.

In questa direzione, rispetto a precedenti esperienze di *record linkage* (ad esempio quello menzionato di Lotti e Marin), il lavoro si distingue per una maggiore copertura delle imprese, inclusi soggetti diversi dalle società di capitale, e per l'utilizzo dell'archivio delle Unità locali (ASIA UL), che migliora sia l'identificazione delle imprese brevettanti sia la localizzazione territoriale dei brevetti. Inoltre, per favorire il collegamento con i brevetti, nel registro longitudinale sono stati mantenuti i record delle imprese presenti in più anni, così da gestire variazioni della ragione sociale dovute a errori, cambi di denominazione od operazioni societarie, elementi identificati come critici nelle esperienze di *record linkage* simili.

Classificazione delle imprese con domande di brevetto

Di seguito vengono mostrati i risultati della classificazione realizzata con il processo di integrazione dei dati precedentemente descritto. Sono state definite le coppie di record corrispondenti ai *match* e *non match*. L'approccio deterministico basato su soglie offre un metodo di classificazione semplice e intuitivo, ma non esente dalla generazione di errori di classificazione rilevanti.

Il principale problema del *record linkage* consistente nel trade-off tra le due usuali tipologie di errore cioè i falsi positivi e i falsi negativi, è stato affrontato rendendo i criteri di abbinamento stringenti per ridurre i falsi positivi aumentando i mancanti abbinamenti, e intervenendo manualmente su quest'ultimi.

Rispetto alla popolazione di imprese italiane selezionate tra il 2015 e il 2023 da EPO, pari a 11.892 imprese, dalle analisi testuali e grazie alle trasformazioni fonetiche descritte in precedenza sono stati individuati doppioni di imprese definite con stringhe diverse. La reale popolazione di imprese individuata ammonta quindi a 9.237 (tabella 2). Nella tabella sono illustrati i risultati dell'assegnazione dei potenziali abbinamenti nelle classi *match e no-match*. Le imprese con codice univoco nei registri statistici Istat identificate come imprese con brevetto EPO rappresentano il 91,5% del totale imprese brevettanti in Italia nel periodo 2015-2023. Mentre le domande di brevetto delle imprese associate a EPO sono pari a 32.133 con una copertura del 95% de totale brevetti.

Rispetto alla quota di *non match* ottenuti con il *record linkage* deterministico processato con il metodo delle soglie, sono state abbinate ulteriori 532 imprese con brevetti tramite procedura manuale.

Tabella 2. Classificazione delle imprese con brevetto EPO tra match e non match (2015-2023)

ANNO	MATCH			NO-MATCH		
	SCORE Block Index	Score Sorted Neighbourhood	Abbinamento manuale	Falsi positiv Block Index	Falsi positivi Sorted Neighbourhood Index	Non match
2015	578	243	48	-	23	-
2016	573	283	48	-	26	-
2017	609	285	76	-	38	-
2018	650	295	65	-	35	-
2019	566	255	66	-	27	-
2020	550	251	80	-	14	-
2021	690	265	37	2	40	148
2022	667	258	66	2	24	195
2023	659	246	45	22	21	166
Totale	**5.542**	**2.381**	**531**	**26**	**248**	**509**
		8.454			**783**	

Fonte: elaborazione dell'Autore su dati EPO e Registri statistici Istat

Per valutare l'efficacia e qualità del processo di produzione di informazioni sulle imprese brevettanti, sono stati calcolati gli indicatori sugli errori di mancato abbinamento e di falso abbinamento, in particolare le metriche Precision (proporzione di coppie di record classificate come corrispondenti che sono effettivamente corrispondenti) e Recall (la proporzione di tutte le corrispondenze effettive che sono identificate come corrispondenti).

Per il calcolo delle due metriche è stato preso a riferimento come gold standard il sottoinsieme di abbinamenti individuato come corretto tramite un processo circolare e ricorsivo di controlli visivi e armonizzazione delle variabili di matching, effettuati sull'output dell'abbinamento con score deterministico pari a 4 (somma dei singoli score delle quattro variabili di abbinamento pari a 1), in entrambi gli step con algoritmo di indicizzazione Block e Neighbourhood. Il valore della Precisione è pari a 0,97 (Veri positivi/Veri positivi+Falsi positivi), mentre il valore della Recall è pari a 0,94 (Veri positivi/Veri positivi+Falsi negativi).

4. Risultati

Nel 2023 sono state pubblicate circa 4.000 domande di brevetto EPO dalle imprese italiane (figura 1). Tra il 2015 e il 2023 il numero di brevetti medi per impresa è pari a 3,9 brevetti con un massimo di 825 brevetti.

L'aumento del numero dei brevetti del 2023 è trainato dal contributo delle imprese, più che da altri gruppi di inventori accademici, istituzionali o individuali. Rispetto al 2015 l'incremento percentuale delle imprese con almeno un brevetto è stato del 48,9, e nello stesso periodo si registra una crescita significativa anche del numero di brevetti depositati dalle imprese, ma con un tasso di variazione maggiore pari al 59,9%. Tale differenza indica una tendenza verso nuove invenzioni più che un aumento di nuove imprese brevettanti.

Il 42,6% delle imprese italiane sono 'One-shot' (Lotti e Nobile 2025) cioè brevettano una sola volta, e questo gruppo è in crescita (figura 2). La maggior parte di queste (85,2%) sono 'Single-patent firms', con un solo brevetto nel periodo considerato, seguendo lo stesso trend. L'aumento di attività innovative isolate evidenzia la difficoltà di trasformare la creatività in produzione stabile e di competere con imprese a innovazione consolidata. Per contro, solo il 7,8% delle imprese, definite 'Incumbent'[2] e dominanti nel mercato dell'innovazione, mostra una continuità significativa nella brevettazione.

La relazione tra imprese a brevettazione persistente e occasionale viene analizzata tramite il co-patenting, ossia i brevetti depositati da due o più entità indipendenti, pratica che rafforza gli ecosistemi di innovazione favorendo collaborazioni tra imprese, anche piccole, e università o organismi di ricerca pubblici, in linea con le opportunità offerte dall'open innovation.

2 Imprese che dominano il mercato delle innovazioni con brevetti per almeno 5 anni e fino a coprire ogni anno del periodo considerato.

Figura 1. Imprese e brevetti 2015-2023

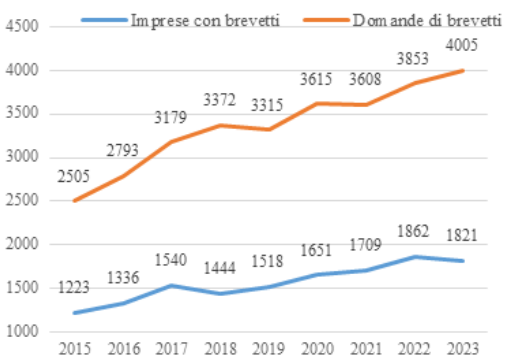
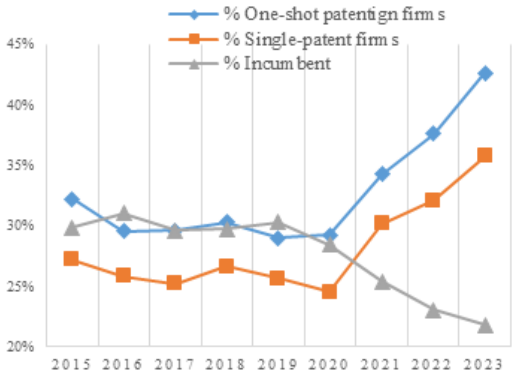


Figura 2. Persistenza nei brevetti imprese 2015-2023



Fonte: elaborazione dell'Autore su dati EPO e Registri statisti Istat

Tabella 3. Domande di brevetto delle imprese per classe di addetti (2015-2023)

Classe di addetti	2015	2016	2017	2018	2019	2020	2021	2022	2023
0-9	21%	22%	21%	26%	25%	25%	26%	27%	30%
10-19	10%	10%	10%	10%	10%	10%	10%	11%	11%
20-49	17%	16%	16%	15%	15%	16%	18%	17%	15%
50-99	13%	16%	16%	14%	14%	13%	13%	13%	13%
100-250	18%	18%	17%	15%	16%	18%	15%	14%	15%
250+	20%	19%	19%	20%	20%	18%	17%	18%	17%
TOT	100%	100%	100%	100%	100%	100%	100%	100%	100%

Fonte: elaborazione dell'Autore su dati EPO e Registri statisti Istat

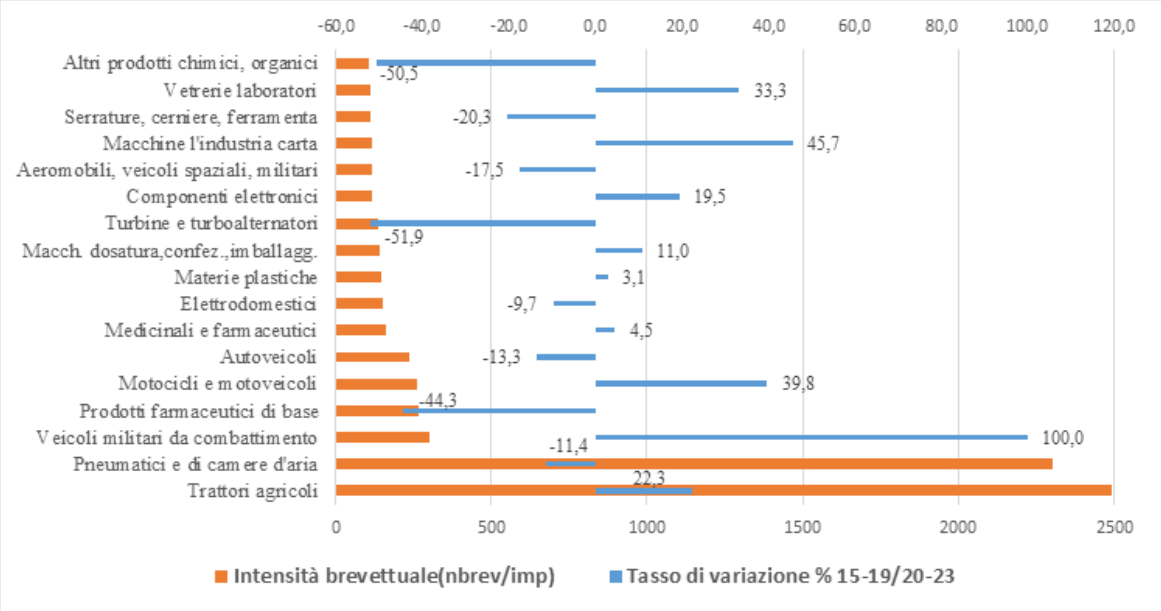
Tra 2015 e 2023 solo il 6,1% delle imprese brevetta in collaborazione; tuttavia, la propensione al co-patenting cresce con la persistenza brevettuale: riguarda il 16,1% delle imprese incumbent (almeno 5 anni di brevetti) e il 36,5% di quelle che depositano brevetti ogni anno.

La dimensione aziendale delle imprese che brevettano permette di comprendere meglio la struttura brevettuale italiana. Quasi la metà delle domande di brevetti depositate tra il 2015 e il 2023 appartiene ad imprese con 250 addetti e oltre (42,9%). Dal punto di vista temporale, le variazioni più rilevanti nella composizione per classe dimensionale riguardano i brevetti delle piccole imprese con meno di dieci addetti che si affidano alla tutela brevettuale pari al 21% nel 2015 e al 30% del 2023 (tabella 3). Rimane costante il peso delle imprese tra i 10 e 100 addetti, mentre diminuisce la quota delle grandi, in particolare tra 100-249 addetti (dal 18% al 15%).

L'integrazione dei brevetti a livello di micro-dato con le imprese e le relative caratteristiche di struttura permette di analizzare i settori di attività economica nei quali le invenzioni brevettate sono diffuse. Osservando le attività a più alta intensità brevettuale (rapporto tra brevetti e imprese per attività economica), oltre che per quantità, i brevetti sono diffusi principalmente nel manifatturiero e, in ordine di intensità, nello specifico nella fabbricazione di: trattori agricoli (tasso di variazione tra il 2015-2019 e 2020-2023 pari al 22,3%), pneumatici e camere d'aria (-11,4), veicoli militari da combattimento (100,0, tenendo conto in questo caso di pochi brevetti, sotto le 10 unità), prodotti farmaceutici di base (-44,3), motocicli e motoveicoli (39,8) (figura 3). In termini assoluti l'attività economica (5 digit) con il maggior numero di brevetti tra il 2015 e il 2023 Fabbricazione di macchine automatiche per la dosatura, la confezione e per l'imballaggio (circa 600 brevetti).

L'integrazione dei micro-dati consente di analizzare non solo i settori più innovativi, ma anche le performance economiche delle imprese brevettanti attraverso gli indicatori dei registri statistici Istat. Nel periodo 2015-2023, le imprese con domande di brevetto presentano una produttività del lavoro nettamente superiore rispetto a quelle senza

Figura 3. Intensità brevettuale (brevetti/imprese) 2023 e tasso di variazione 2015-2019 - 2020/2023 (%)



Fonte: elaborazione dell'Autore su dati EPO e registri statisti Istat

brevetti (circa 116 mila euro contro 49 mila euro per addetto). I dati mostrano inoltre che queste imprese hanno livelli iniziali di produttività mediamente maggiori e differenze particolarmente accentuate per dimensione aziendale, con incrementi del 31,6% tra le micro-imprese (meno di 10 addetti) e del 23,9% tra le grandi imprese (oltre 250 addetti).

Poiché i brevetti coprono principalmente invenzioni tecniche, rappresentano una fonte naturale di informazioni relative ai cambiamenti tecnologici, e spesso l'unica fonte affidabile per lo studio di tecnologie emergenti non ancora diffuse (OECD 2009). L'ampiezza tecnologica dei brevetti (*patent scope*) nel portafoglio di un'impresa è spesso associata al valore tecnologico ed economico dell'impresa stessa. I brevetti con ampiezza tecnologica maggiore hanno più valore poichè incorporano più tecnologie nella stessa classe di prodotto della classificazione IPC[3] (Classificazione internazionale dei brevetti).

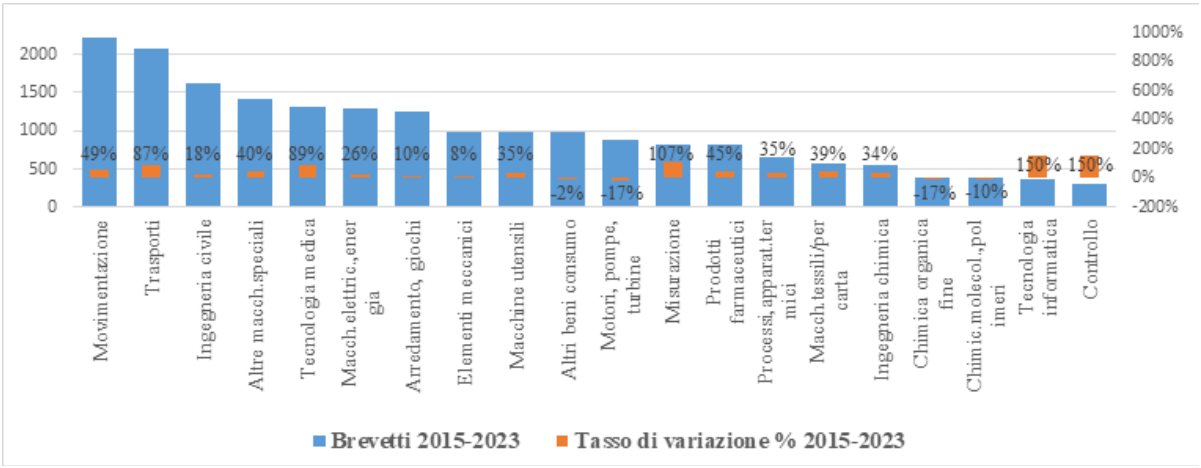
Il trend di innovazione nel periodo considerato indica campi tecnologici diversi da quelli rilevanti e stabili collocati nei primi 20 per numero di brevetti tra il 2015 e il 2023 (figura 4). I tassi di variazione più alti sono in ordine nei Processi di comunicazione di base, nella Micro-nanotecnologia, nei Semiconduttori (campo tecnologico con il salto maggiore dalla 32° alla 18° posizione) e Tecnologia ambientale e Comunicazione digitale: questi campi non sono tra i primi 20 per contributo al portafoglio di brevetti delle imprese italiane, piuttosto agli ultimi posti per consistenza (quota sui brevetti italiani tra 0,1 e 0,7%). L'analisi congiunta delle due dimensioni consistenza e incremento brevettuale suggerisce l'avvio di un percorso di crescita in questi settori, ma una scarsa consistenza innovativa proprio nelle aree *core* dell'intelligenza artificiale in crescita in altri Paesi, poiché semiconduttori e tecnologie micro-nano elettroniche rappresentano la spina dorsale dell'intelligenza artificiale (hardware) e i brevetti sulla Comunicazione digitale, Processi di comunicazione di base e Tecnologia informatica l'altra medaglia dell'innovazione (software).

Trend negativi per i campi della chimica: Chimica alimentare (-31%), Chimica organica fine (-17%) e Chimica macromolecolare, polimeri (-10%), e Motori, pompe, turbine (-17%).

L'analisi dei dati brevettuali evidenzia il settore dei semiconduttori come ambito tecnologico emergente e strategico, pur non rientrando tra i primi 20 settori per numero assoluto di brevetti nel periodo 2015-2023. La progettazione e il design dei semiconduttori mostrano infatti una forte crescita innovativa e sono stati individuati dalla Commissione europea come settore ad alta intensità di R&S e strategico (Higuera 2025). In questo contesto, l'Italia ha recepito le

3 L'Organizzazione mondiale per la proprietà intellettuale (WIPO) classifica i brevetti in 5 settori tecnologici (Chimica, Ingegneria elettrica, Ingegneria meccanica, Strumentazione, Altri campi) suddivisi in 35 campi tecnologici. È possibile visionare la classificazione Ipc al link: https://www.wipo.int/en/web/classification-ipc.

Figura 4. Primi 20 campi tecnologici per numero di brevetti e tasso di variazione 2015-2023



Fonte: elaborazione dell'Autore su dati EPO e registri statisti Istat

indicazioni del Chips Act attraverso iniziative come il Chips Fund 2022 e il National Chips Act.

Il sistema integrato impresa-brevetto sviluppato nel presente lavoro consente di identificare caratteristiche e profili delle imprese attive nel settore. I semiconduttori registrano il più alto incremento brevettuale tra i 35 campi tecnologici IPC nel periodo considerato (+253%). Circa il 75% delle invenzioni è concentrato nella fabbricazione di computer e prodotti elettronici, seguita dalla fabbricazione di macchinari e apparecchiature e poi ricerca scientifica e sviluppo.

Rispetto al totale delle imprese brevettanti italiane, quelle dei semiconduttori mostrano maggiore continuità innovativa: il 31% è costituito da imprese one-shot contro il 42,6% complessivo. Il settore presenta inoltre una quota elevata di co-patenting (28%), rispetto al 6% delle imprese brevettanti totali, evidenziando l'importanza delle collaborazioni nella ricerca. Dal punto di vista dimensionale emerge una forte polarizzazione, con una maggiore presenza sia di grandi imprese (23% con oltre 250 addetti contro il 17% totale) sia di microimprese.

Alla luce dell'evoluzione del Chips Act europeo 2.0 (giugno 2026), che punta a rafforzare ricerca e innovazione, stimolare la domanda di chip e la formazione di ecosistemi tra produttori e utilizzatori di chip, i risultati suggeriscono due principali linee di intervento: sostenere le imprese che hanno già sperimentato brevetti nel settore ma non mantengono continuità innovativa, spesso per vincoli dimensionali o di accesso agli incentivi; favorire progetti collaborativi tra imprese, istituzioni, aziende *fabless* (design) e *foundries* (produzione). La rilevanza delle partnership e l'elevata intensità innovativa del settore confermano infatti la necessità di ecosistemi capaci di integrare ricerca, progettazione, domanda industriale e produzione di chip, quest'ultima sostenibile solo con volumi produttivi elevati, che consentono di beneficiare di economie di scala, e di legami forti tra domanda e innovazione.

La classificazione per settori tecnologici integrata con i micro-dati sulle imprese permette di analizzare quindi settori tecnologici emergenti. Visto il crescente interesse generale per l'intelligenza artificiale e la ricchezza unica dei dati brevettuali per misurare le invenzioni tecnologiche, di seguito viene mostrata l'evoluzione dell'intelligenza artificiale in Italia tramite le invenzioni delle domande di brevetto pubblicate dal 2015 al 2023 (OECD 2025)[4]. Il numero di brevetti in AI è cresciuto dal 2015 al 2023 con un tasso di variazione maggiore (117%) rispetto allo stock totale di brevetti delle imprese italiane (60%), registrando 1145 domande di brevetti cumulate, in particolare nel 2022 e 2023, delle quali il 43,4% dal 2021 al 2023.

Conclusioni

Gli output illustrati nel presente lavoro forniscono analisi su specifiche caratteristiche del sistema produttivo italiano sotto il profilo della propensione a innovare tramite i brevetti.

L'utilizzo dei brevetti come indicatore dell'innovazione presenta sia punti di forza sia

4 La metodologia adottata per la classificazione dei brevetti in intelligenza artificiale prende spunto dalla nuova tassonomia pubblicata dall'OECD (Settembre 2025), *Identifying emerging AI technologies using patent data*, utilizzando i codici CPC definiti nello studio.

limiti. La letteratura economica riconosce nei dati brevettuali la misura osservabile più sistematica dell'attività inventiva e un indicatore strettamente correlato agli investimenti in R&S (Griliches 1998). Tuttavia, i brevetti misurano l'attività inventiva formalizzata e legalmente tutelabile, non l'innovazione in senso pienamente schumpeteriano.

La loro capacità informativa è limitata dal fatto che la propensione a brevettare varia tra settori e dimensioni d'impresa; molte innovazioni, soprattutto nei servizi e nelle PMI, vengono protette tramite segreto industriale o vantaggi competitivi temporanei. Inoltre, il valore dei brevetti è distribuito in modo fortemente asimmetrico e può riflettere anche strategie competitive (*strategic patenting*), oltre che reale attività inventiva.

I brevetti rappresentano quindi un indicatore affidabile della capacità tecnologica e della produzione di conoscenza codificata, ma una misura imperfetta della propensione innovativa complessiva. Il loro utilizzo è particolarmente appropriato nei settori ad alta intensità di ricerca e sviluppo, mentre richiede cautela nei contesti caratterizzati da innovazione tacita, organizzativa o di processo. Per questo motivo dovrebbero essere integrati con altre informazioni, quali investimenti in R&S, produttività e introduzione di nuovi prodotti.

Nonostante tali limiti, i dati brevettuali svolgono un ruolo centrale nelle politiche pubbliche per l'innovazione, poiché consentono di analizzare in modo sistematico la produzione di conoscenza tecnologica, l'accumulazione di asset intangibili e le traiettorie settoriali dell'innovazione grazie alla loro natura standardizzata, territoriale e temporale.

Un ambito di applicazione riguarda le politiche fiscali per la valorizzazione della proprietà intellettuale, tra cui spiccano i regimi del Patent Box e Brevetti+, come rimodulati negli ultimi anni. In questi schemi, i dati sui brevetti possono essere utilizzati per identificare gli asset eleggibili, collegare i redditi agevolati alle attività di ricerca e sviluppo sottostanti e prevenire comportamenti opportunistici. L'informazione brevettuale consente inoltre di valutare ex post l'efficacia di tali strumenti, analizzando se l'incentivo fiscale influenza aumenti dell'attività inventiva domestica, la localizzazione degli asset intangibili o fenomeni di concentrazione settoriale.

Un altro filone interessante è quello delle politiche di sostegno diretto alla R&S e all'innovazione, come sussidi, grant competitivi e programmi di collaborazione pubblico-privato. In questo contesto, i brevetti vengono impiegati sia come criterio di selezione (per valutare le capacità innovative pregresse dei beneficiari) sia come indicatore di output per misurare l'impatto degli interventi. L'analisi delle citazioni e del livello tecnologico dei brevetti permette inoltre di cogliere effetti qualitativi, come l'originalità o la rilevanza delle invenzioni generate.

I dati brevettuali sono centrali anche nelle politiche industriali e tecnologiche, dove aiutano a identificare specializzazioni produttive, vantaggi tecnologici comparati e aree strategiche emergenti quale, naturalmente il settore emergente dell'intelligenza artificiale, le tecnologie verdi, digitali e biomedicali, o l'area dei materiali avanzati come i semiconduttori (si veda il Chips Act italiano), emersi nei risultati delle specializzazioni tecnologiche di questo lavoro.

In questo senso, i brevetti supportano la definizione delle priorità di policy e il monitoraggio dell'evoluzione tecnologica nel tempo, offrendo una base empirica per interventi mirati a livello settoriale o territoriale.

Un ulteriore ambito di applicazione empirica riguarda le politiche per il trasferimento tecnologico. L'informazione brevettuale consente di valutare la capacità di università ed enti pubblici di ricerca di trasformare conoscenza scientifica in applicazioni industriali. In questo caso i brevetti incarnano indicazioni sulle relazioni per l'innovazione nei casi di brevetti sviluppati in collaborazione tra più soggetti, imprese con imprese o imprese e settori della ricerca istituzionale.

In generale si evidenzia come i dati sui brevetti siano uno strumento trasversale alle politiche per l'innovazione, ma anche come il loro utilizzo richieda cautela. Tuttavia, se integrati con altre fonti informative, essi rappresentano una base empirica solida per politiche più informate, mirate e valutabili.

La disponibilità di informazioni a livello di micro-dato e la granularità delle informazioni prodotte in questo lavoro permette di disporre di dati, ad esempio per andare oltre il criterio degli interventi pubblici *one size fits all.* L'individuazione e selezione dei beneficiari di specifici interventi pubblici, come le imprese che brevettano *one-shot*, può consentire interventi mirati, risparmi di risorse pubbliche e sfavorire effetti di monopolio e aumenti del divario tecnologico e di innovazione tra grandi e piccole imprese. Le politiche pubbliche possono quindi valutare l'efficacia degli incentivi brevettuali, modulare gli incentivi, intervenire con licenze obbligatorie, definire priorità tecnologiche personalizzabili ad esempio per territorio o per specializzazione settoriale.

Bibliografia

Athey S., Chetty R., Imbens G. (2020), *Using Experiments to Correct for Selection in Observational Studies*, Papers 2006.09676, arXiv.org, revised May 2025

Azzone G. (2018), Big data and public policies: Opportunities and challenges, *Statistics & Probability Letters*, 136, pp.116-120

Barcaroli G., Nurra A., Salamone S., Scannapieco M., Scarnò M., Summa D. (2015), Internet as Data Source in the Istat Survey on ICT in Enterprises, *Austrian Journal of Statistics*, 44, n.2, pp.31-43

Boix-Domènech R., Sforzi F., Galletto V., Capone F. (2025), Born to innovate, not to produce: three decades of compared change, evolution and innovation in the dynamic Marshallian Industrial Districts of Spain and Italy, *European Planning Studies*, 33, n.6, pp.899-916

Card D., Chetty R., Feldstein, M.S., Saez E. (2010), Expanding Access to Administrative Data for Research in the United States, *American Economic Association, Ten Years and Beyond: Economists Answer NSF's Call for Long-Term Research Agendas* <doi.org/10.2139/ssrn.1888586>

Christen P. (2012), *Data matching: concepts and techniques for record linkage, entity resolution, and duplicate detection*, Berlin-Heidelberg, Springer

Connelly R., Playford C.J., Gayle V., Dibben C. (2016), The role of administrative data in the big data revolution in social science research, *Social Science Research*, 59, pp.1-12

Crato N., Paruolo P. (eds.) (2019), *Data-Driven Policy Impact Evaluation*. *How Access to Microdata is Transforming Policy Design,* Cham*,* Springer

EPO (2025), *Standards and the European Patent System: Insights from a New EPO Dataset Linking Patents and Standards, with Early Perspectives into SEP Litigation under the Unified Patent Court*, Munich, European Patent Office

Griliches Z. (1998), Patent Statistics as Economic Indicators: A Survey, in *R&D and Productivity: The Econometric Evidence*, Chicago and London, University of Chicago Press, pp.287-343

Higuera A.G. (2025), *What if the EU ran on microelectronics?*, European Parliamentary Research Service, October

Jabine T.B., Scheuren F.J. (1986), Record linkages for statistical purposes: methodological issues, *Journal of Official Statistics*, 2, pp.255-277

Jaro M.A. (1989), Advances in Record-Linkage Methodology as Applied to Matching the 1985 Census of Tampa, Florida, *Journal of the American Statistical Association*, 84, n.406, pp.414-420

Lotti F., Nobile C. (2025), *The geography of innovation: patent insights into Europe's green and digital transitions*, Questioni di Economia e Finanza (Occasional Papers) n.945, Roma, Banca d'Italia

Lotti F., Marin G. (2013), *Matching of PATSTAT applications to AIDA firms: discussion of the methodology and results*, Questioni di Economia e Finanza (Occasional Papers) n.166, Roma, Banca d'Italia

Lotti F., Schivardi F. (2005), Cross country differences in patent propensity: A firm-level investigation, *Giornale degli Economisti,* 64, n.4, pp.469-502

OECD (2025), *Identifying emerging AI technologies using patent data: A semi-automated approach*, Paris, OECD Publishing

OECD, Eurostat (2018), *Oslo Manual 2018: Guidelines for Collecting, Reporting and Using Data on Innovation, 4th Edition, The Measurement of Scientific, Technological and Innovation Activities*, Paris, OECD Publishing / Luxembourg, Eurostat

OECD (2015), *Frascati Manual 2015: Guidelines for Collecting and Reporting Data on Research and Experimental Development, The Measurement of Scientific, Technological and Innovation Activities*, Paris, OECD Publishing

OECD (2009), *Patent Statistics Manual*, Paris, OECD Publishing

Pakes A., Griliches Z. (1984), *Patents and R&D at the Firm Level: A First Look*, in Griliches Z. (ed.), *R&D, Patents, and Productivity*, Chicago, University of Chicago Press for the National Bureau of Economic Research, pp.55-72

Pavitt K. (1984), Sectoral patterns of technical change: Towards a taxonomy and a theory, *Research Policy*, 13, n.6, pp.343-373

Scanu M. (a cura di) (2013), *Metodi statistici per il record linkage*, Roma, Istat

Schmookler J. (1966), *Invention and Economic Growth*, Cambridge (MA), Harvard University Press

Shlomo N. (2019), Overview of Data Linkage Methods for Policy Design and Evaluation, in Crato N., Paruolo P. (eds.), *Data-Driven Policy Impact Evaluation*, Springer, Cham, pp.47-65

Squicciarini M., Dernis H., Criscuolo C. (2013), *Measuring patent quality: Indicators of technological and economic value*, OECD Science, Technology and Industry Working Papers 2013/03, Paris, OECD Publishing

Thoma G., Torrisi S., Gambardella A., Guellec D., Hall B.H., Harhoff D. (2010), *Harmonizing and Combining Large Datasets. An Application to Firm-Level Patent and Accounting Data*, NBER Working Papers n.15851, Cambridge (MA), National Bureau of Economic Research

Trajtenberg M. (1990), A Penny for Your Quotes: Patent Citations and the Value of Innovations, *The RAND Journal of Economics*, 21, n.1, pp.172-187

von Lucke J., Frank S. (2025), Potentials of administrative informatics for the analysis of policymaking: notes on the integration of administrative informatics into the policy cycle, *Data & Policy*, 7:e47 <doi:10.1017/dap.2025.10010>

Winkler W.E. (1990), String Comparator Metrics and Enhanced Decision Rules in the Fellegi-Sunter Model of Record Linkage, in *Proceedings of the Survey Research Methods Section*, Alexandria (VA), American Statistical Association, pp.354-359

Sergio Salamone
sesalamo@istat.it
Ricercatore Istat, si è occupato di ricerca nel settore dell'economia della conoscenza, della proprietà intellettuale e digitalizzazione nelle imprese e istituzioni. È autore di pubblicazioni istituzionali e scientifiche su digitalizzazione e sviluppo regionale.

Dagli aggregati ai microdati

Spesa sanitaria privata e incentivi fiscali in Italia

Clara Andresciani
Sapienza Università di Roma

Nel dibattito pubblico, la spesa sanitaria privata, l'assicurazione sanitaria volontaria e gli incentivi fiscali sono sempre più considerati strumenti per migliorare la capacità di risposta e decongestionare i sistemi sanitari pubblici. La verifica empirica di tale ipotesi richiede tuttavia dati coerenti e sufficientemente disaggregati. Con riferimento all'Italia, questo lavoro analizza le principali fonti istituzionali sulla spesa sanitaria privata, la copertura assicurativa e le detrazioni fiscali, evidenziandone alcuni limiti: la scarsa comparabilità delle stime, la difficoltà di distinguere la spesa out-of-pocket da quella assicurativa e la limitata disponibilità di dati per analisi distributive. Il Sistema Tessera sanitaria rappresenta un'infrastruttura ancora sottoutilizzata che, attraverso un accesso sicuro ai dati individuali per finalità di ricerca, potrebbe consentire una misurazione più rigorosa della spesa privata e una migliore valutazione degli incentivi fiscali e dei meccanismi di finanziamento privato.

From aggregate data to microdata. Private healthcare expenditure and tax incentives in Italy

Private healthcare expenditure, voluntary health insurance, and tax incentives are increasingly considered instruments to improve responsiveness and relieve pressure on public health systems. Testing this hypothesis requires data that are conceptually coherent and sufficiently granular. Focusing on Italy, this paper analyses the main institutional sources on private healthcare expenditure, insurance coverage, and tax deductions, highlighting key limitations: the lack of comparability across estimates, the difficulty of distinguishing out-of-pocket payments from insurance-related expenditure, and the limited availability of data for distributional analysis. The Sistema Tessera sanitaria represents an underexploited administrative infrastructure that, through secure research access to individual-level data, could enable a more rigorous measurement of private expenditure and a better evaluation of fiscal incentives and private financing mechanisms.

DOI: 10.53223/Sinappsi_2026-02-09

Citazione
Andresciani C. (2026), Dagli aggregati ai microdati. Spesa sanitaria privata e incentivi fiscali in Italia, *Sinappsi*, XVI, n.2, pp.135-144

Parole chiave
Cittadini
Dati amministrativi
Sanità

Keywords
Citizens
Administrative data
Health service

Introduzione

Negli ultimi anni, il ruolo del finanziamento sanitario privato nei sistemi sanitari europei ha ricevuto crescente attenzione nel dibattito di policy e nella letteratura comparata sui sistemi sanitari. In diversi Paesi europei, la crescita della spesa *out-of-pocket* (OOP) e, in misura eterogenea, della copertura assicurativa sanitaria volontaria si è sviluppata parallelamente alle persistenti pressioni sui sistemi pubblici legate all'invecchiamento della popolazione, ai vincoli fiscali, all'aumento della domanda di cura e ai problemi di accessibilità,

inclusi tempi di attesa e carenze di personale sanitario (OECD 2024). In questo quadro, il ricorso all'assicurazione sanitaria volontaria e ad altre forme di finanziamento privato è stato spesso interpretato come una risposta alle difficoltà di accesso e ai limiti di capacità dei sistemi pubblici, sebbene con rilevanti differenze istituzionali tra paesi europei (Purcel *et al*. 2023). Un maggiore coinvolgimento del settore privato è pertanto talvolta discusso come possibile strumento complementare in grado di ampliare la capacità di risposta dei sistemi sanitari e offrire canali alternativi di accesso alle cure.

Anche nel caso italiano, la spesa sanitaria privata, la copertura assicurativa volontaria e gli incentivi fiscali associati sono frequentemente richiamati nel dibattito pubblico e istituzionale come potenziali mezzi per migliorare l'accesso ai servizi e ridurre la pressione sul Servizio Sanitario Nazionale (SSN). Questa argomentazione richiama il cosiddetto 'effetto decongestione', secondo cui, se una quota crescente di individui ricorre al settore privato, parte della domanda sanitaria si sposterebbe fuori dal sistema pubblico, liberando capacità e risorse che potrebbero tradursi in un miglioramento dell'accesso per i pazienti che continuano a utilizzare il SSN. Sebbene tale meccanismo sia spesso evocato nel dibattito sulle interazioni tra pubblico e privato nei sistemi sanitari misti, le condizioni teoriche ed empiriche necessarie affinché esso si realizzi restano oggetto di dibattito e non trovano evidenza empirica univoca nella letteratura esistente (Besley *et al.* 1999; Propper 2000; Propper and Green 2001).

La possibilità di valutare empiricamente queste dinamiche e le relative implicazioni dipende tuttavia dalla disponibilità di dati affidabili, comparabili e sufficientemente disaggregati sulle diverse componenti del finanziamento sanitario privato.

Nel caso italiano, questo presupposto risulta solo parzialmente soddisfatto poiché il panorama informativo resta frammentato e caratterizzato da discontinuità che ostacolano il monitoraggio e la valutazione delle dinamiche del finanziamento privato e dei suoi effetti su accesso ed equità.

Innanzitutto, le stime della spesa sanitaria privata divergono tra fonti istituzionali perché poggiano su architetture contabili e classificazioni differenti – ad esempio COICOP[1]/Conti nazionali rispetto a SHA/ICHA[2] – che implicano perimetri, definizioni operative e criteri di imputazione non pienamente sovrapponibili. A ciò si aggiungono perimetri di riferimento diversi (territorio economico versus popolazione residente) e, soprattutto, scelte di confine non univoche su ciò che debba essere considerato 'privato': anche aspetti apparentemente tecnici, come il trattamento dei ticket o dei rimborsi e delle agevolazioni fiscali, possono incidere in modo significativo sui livelli stimati e sulla comparabilità nel tempo e nello spazio. La misurazione della spesa privata è ulteriormente ostacolata dall'assenza di informazioni sistematiche a livello individuale, condizione che vincola la possibilità di condurre analisi distributive e orientate all'equità: senza microdati non è possibile osservare con continuità chi sostiene la spesa, per quali prestazioni e con quale intensità, né ricostruire come tali profili varino per reddito, età, condizione occupazionale o area di residenza. Questa lacuna rende più difficile anche distinguere empiricamente tra spesa OOP, compartecipazioni al SSN e spesa intermediata o rimborsata da coperture assicurative, con ricadute dirette sulla solidità della base empirica disponibile.

Inoltre, la copertura assicurativa sanitaria volontaria rimane difficile da quantificare in termini che siano, al contempo, comparabili sul piano internazionale e utilizzabili sul piano nazionale per analisi empiriche robuste. Le criticità riguardano sia la tassonomia utilizzata da differenti organismi internazionali sia la disponibilità di dati granulari che consentano di studiarne la distribuzione per caratteristiche socioeconomiche e territoriali e, soprattutto, la sua interazione con l'uso dei servizi e con la spesa privata. Ne deriva un paradosso informativo: la copertura assicurativa sanitaria è frequentemente evocata come una potenziale soluzione alle criticità di accesso e ai tempi di attesa, nonché come uno strumento di alleggerimento della pressione sul SSN, ma rimane solo parzialmente osservabile con gli strumenti correnti, limitando la possibilità di valutare in modo rigoroso i suoi effetti.

Infine, le detrazioni fiscali per spese sanitarie – leva di policy rilevante sia per l'ammontare delle risorse mobilitate sia per i potenziali effetti distributivi – non possono essere valutate con pieno rigore utilizzando i dati oggi disponibili, poiché le statistiche esistenti non consentono di distinguere componenti di spesa sostanzialmente eterogenee. In assenza di

1 COICOP (*Classification of Individual Consumption by Purpose)* è la classificazione dei consumi individuali secondo lo scopo). È uno standard internazionale creato dalle Nazioni Unite per suddividere le spese di beni e servizi in gruppi chiari e ordinati.

2 SHA e ICHA sono gli acronimi usati dall'Istat per indicare il *System of Health Accounts* (Sistema dei conti della sanità) e l'*International Classification for Health Accounts* (Classificazione internazionale dei conti sanitari). Servono a misurare e confrontare la spesa sanitaria pubblica e privata in modo uguale in tutta Europa. Fonte: https://www.istat.it/scheda-qualita/sistema-dei-conti-della-sanita/.

una separazione informativa tra compartecipazioni al SSN (ticket), spesa pienamente privata e spesa intermediata o rimborsata da assicurazioni, risulta difficile stimare chi benefici effettivamente degli incentivi, in quale misura e con quali implicazioni per equità, accesso e protezione finanziaria.

All'interno di questo quadro, il Sistema Tessera sanitaria (STS) – gestito dal Ministero dell'Economia e delle finanze tramite la Ragioneria generale dello Stato – rappresenta un'opportunità cruciale e, per molti aspetti, ancora non pienamente valorizzata. In particolare, tale sistema raccoglie informazioni amministrative sulle transazioni sanitarie trasmesse in via elettronica dai fornitori e, in linea di principio, potrebbe costituire un'infrastruttura in grado di ridurre alcune delle principali opacità del finanziamento privato. Se i microdati individuali del sistema fossero accessibili per finalità di ricerca, sarebbe possibile distinguere la spesa OOP da quella rimborsata da assicurazioni, separare le compartecipazioni al SSN dalle prestazioni pienamente private e analizzare con maggiore precisione come spesa privata e incentivi fiscali si distribuiscano tra gruppi di popolazione e territori. Ciò consentirebbe di passare da un monitoraggio prevalentemente aggregato a un'analisi più dettagliata e informativa dei meccanismi sottostanti, inclusi i profili di regressività/progressività e le possibili interazioni tra copertura assicurativa, comportamento di spesa e utilizzo dei servizi.

Alla luce di queste premesse, l'obiettivo del presente lavoro è descrivere e discutere criticamente le principali sfide di misurazione relative alla spesa sanitaria privata, alla copertura assicurativa e alle detrazioni fiscali in Italia, mostrando in che modo tali limiti condizionino il monitoraggio orientato all'equità e la valutazione delle politiche pubbliche. Mettendo in relazione questioni tecniche di misurazione – contabilizzazione e accesso ai microdati – con domande di policy ben definite, come la distribuzione dei benefici fiscali e l'interazione tra copertura assicurativa, spesa privata e accesso ai servizi, il contributo delimita con maggiore chiarezza ciò che le fonti attualmente disponibili consentono di osservare e ciò che, al contrario, non è misurabile con adeguata precisione. In un contesto in cui i sistemi sanitari assumono assetti sempre più ibridi, rafforzare l'infrastruttura informativa diventa dunque una condizione necessaria per interpretare correttamente le dinamiche del finanziamento e le loro implicazioni per l'equità.

Il resto del lavoro è organizzato come segue: nella sezione successiva vengono esaminate le principali lacune e limitazioni nella misurazione della spesa sanitaria privata, della copertura assicurativa e delle detrazioni fiscali in Italia, ricostruendone le implicazioni in termini di comparabilità, disaggregazione e possibilità di analisi empirica; nella sezione conclusiva vengono discusse le implicazioni per la governance dei dati e per la capacità valutativa delle politiche pubbliche, indicando le priorità informative necessarie a sostenere decisioni basate su evidenza in materia di finanziamento sanitario.

1. Lacune e limiti nella misurazione di spesa sanitaria privata, copertura assicurativa sanitaria e detrazioni fiscali

Il progressivo coinvolgimento del settore privato nei sistemi sanitari europei ha alimentato un ampio dibattito sulle possibili implicazioni di tali trasformazioni per l'equità nell'accesso alle cure e per la sostenibilità dei principi universalistici che caratterizzano i modelli pubblici di assistenza sanitaria (World Health Organization 2023). Sebbene l'espansione di alternative private sia spesso giustificata in termini di efficienza, maggiore capacità di risposta e decongestione del canale pubblico, l'impatto effettivo e le conseguenze distributive di questo spostamento non possono essere assunti *ex ante* e restano difficili da quantificare in modo convincente. La letteratura ha infatti evidenziato che, nei sistemi in cui l'accesso al settore pubblico è anche razionato tramite i tempi di attesa, l'allungamento delle liste può alimentare la domanda di copertura assicurativa privata come strumento di *bypass* dell'attesa (Besley *et al.* 1999). In modo coerente, evidenze recenti per l'Italia suggeriscono che la disponibilità di una copertura assicurativa sanitaria sia associata a una maggiore probabilità di ricorrere al canale privato per alcune prestazioni e a una diversa modalità di utilizzo del SSN, con implicazioni potenzialmente non neutrali per l'equità (Brenna 2025).

Alla luce di queste dinamiche, comprendere il ruolo del settore privato richiede di misurarne non solo il peso finanziario complessivo ma anche la composizione e la distribuzione dell'onere finanziario, e dunque le sue implicazioni distributive, tra individui e territori. Non è sufficiente conoscere l'ammontare aggregato della spesa sanitaria privata: è altrettanto necessario osservare, a livello individuale, chi sostiene la spesa, per quali voci (ticket, farmaci, diagnostica), quanta parte sia mediata o rimborsata da assicurazioni o fondi aziendali e come gli incentivi fiscali si intreccino con queste componenti. Tuttavia, in Italia le fonti disponibili differiscono per definizioni, metodi e livelli di aggregazione e, soprattutto, non consentono un'analisi a livello individuale e comparabile delle diverse componenti del finanziamento privato. Ne

deriva un vincolo diretto alla valutazione empirica, poiché diventa difficile stimare in modo robusto quali gruppi beneficino dei canali privati e degli incentivi fiscali e quali, al contrario, risultino maggiormente esposti a oneri e barriere economiche.

Le sezioni seguenti descrivono quindi le principali infrastrutture informative italiane sul finanziamento privato, dapprima rivedendo le fonti utilizzate per misurare la spesa sanitaria privata e, successivamente, esaminando i dati disponibili sulla copertura assicurativa volontaria e sulle detrazioni fiscali.

2. Dati sulla spesa sanitaria privata

Il finanziamento sanitario privato comprende sia la spesa OOP sia i premi versati per l'acquisto di coperture assicurative volontarie. Una sua quantificazione accurata richiede però, preliminarmente, di chiarire i confini concettuali dell'oggetto di misura e di ricostruire come le diverse fonti producano e rendano comparabili le proprie stime.

A differenza della spesa sanitaria pubblica, il cui perimetro e le cui modalità di finanziamento risultano relativamente più leggibili, la spesa sanitaria privata può essere delineata in modi diversi a seconda dell'obiettivo analitico e del framework contabile adottato. Su questo punto, Del Vecchio *et al*. (2024) evidenziano due questioni concettuali tra loro strettamente interrelate.

La prima riguarda il criterio utilizzato per attribuire una determinata componente di spesa al settore privato: se la classificazione debba basarsi sulla natura istituzionale del soggetto che raccoglie le risorse (pubblico o privato) oppure sulle caratteristiche del meccanismo di finanziamento, in particolare sul grado di volontarietà del trasferimento finanziario. In alcuni sistemi sanitari europei, infatti, una quota della spesa formalmente privata deriva da schemi a partecipazione vincolata o fortemente regolati dall'intervento pubblico. È il caso, ad esempio, dell'assicurazione sanitaria obbligatoria in Svizzera o di alcune forme di copertura complementare collettiva previste per specifiche categorie di lavoratori in Francia. In questi casi, pur trattandosi di flussi formalmente classificati come privati, il grado di scelta individuale risulta limitato e la distinzione tra finanziamento pubblico e privato diventa meno netta. Una questione distinta, ma concettualmente affine, riguarda invece le compartecipazioni richieste agli utenti all'interno dei sistemi sanitari pubblici. In Italia, i ticket su prestazioni coperte dal SSN sono sostenuti direttamente dalle famiglie ma si collocano all'interno di un'offerta regolata pubblicamente e rispondono a logiche di compartecipazione al finanziamento del sistema. Analogamente, anche i vantaggi fiscali associati ad alcune spese sanitarie riducono l'onere netto effettivamente sostenuto dalle famiglie. Ne deriva che il perimetro della spesa sanitaria privata può essere costruito in modi differenti: ad esempio includendo o escludendo i ticket oppure considerando gli effetti delle agevolazioni fiscali quando l'obiettivo sia stimare il costo privato effettivamente sostenuto.

La seconda questione riguarda quali beni e servizi includere. I confini di ciò che si qualifica come 'Sanità' risultano oggi meno netti, anche in ragione dell'evoluzione dei bisogni di salute, delle trasformazioni demografiche che spingono dalla cura acuta alla gestione della cronicità e della crescente intersezione tra assistenza sanitaria e assistenza sociale. Anche a fronte di definizioni operative relativamente consolidate, permane un margine di discrezionalità nel decidere se includere categorie specifiche – come integratori alimentari, procedure puramente estetiche o forme di assistenza di lungo periodo con una componente sanitaria – che, a seconda delle scelte adottate, possono incidere in modo non trascurabile sulle stime complessive.

Queste differenze concettuali si riflettono nelle principali fonti italiane. I Conti nazionali includono soltanto il consumo finale delle famiglie nel territorio economico e seguono tipicamente la classificazione COICOP, standard internazionale che ripartisce i consumi individuali 'per finalità' di spesa. Seguendo questa metodologia, la spesa delle famiglie viene attribuita in base al tipo di bene/servizio acquistato e allo scopo d'uso (ad esempio, sanità, istruzione, trasporti) e la sezione 'Sanità' include voci quali prodotti farmaceutici e altri prodotti medici, servizi ambulatoriali e servizi ospedalieri.

Le stime basate su SHA, invece, si riferiscono alle famiglie residenti e includono, oltre alla spesa OOP diretta, altre forme di finanziamento volontario, come assicurazioni private e schemi offerti dai datori di lavoro al di fuori dei meccanismi contributivi pubblici. Sviluppato da OECD, Eurostat e WHO, lo SHA rappresenta lo standard internazionale per la contabilità della spesa sanitaria e consente di confrontarne il livello e la composizione tra paesi e nel tempo. Diversamente da un approccio centrato sul consumo finale, tale approccio ricostruisce la spesa sanitaria distinguendo sistematicamente cosa viene acquistato (funzioni e servizi), chi eroga le prestazioni (fornitori) e attraverso quali schemi di finanziamento.

La tabella 1 sintetizza le principali fonti istituzionali utilizzate per misurare la spesa sanitaria privata in Italia, indicando approcci di aggregazione e classificazione, anni coperti, stime più recenti e livello di disaggregazione geografica. Il dettaglio

Tabella 1. Principali fonti istituzionali per la misurazione della spesa sanitaria privata in Italia

Fonte	Aggregati e metodi di riferimento	Classificazione	Stima nazionale (mld EUR, ultimo anno)	Anni disponibili	Revisionato nel tempo	Dettaglio geografico
ISTAT	Spesa sanitaria rilevata dall'Indagine sui consumi delle famiglie (sostenuta in Italia e all'estero)	COICOP 2018	35,9	1997-2022	No	Nazionale e regionale
ISTAT	Consumo finale di famiglie residenti e non residenti nel territorio economico, stimato secondo il Sistema dei conti nazionali (SEC 2010), integrando dati amministrativi e campionari	COICOP 1999	43,0	1995-2023	Sì	Nazionale
ISTAT	Spesa sanitaria sostenuta dalla popolazione residente (spesa diretta delle famiglie e schemi volontari) stimata secondo il System of Health Accounts (SHA 2011)	SHA 2011	45,9	2012-2023	Sì	Nazionale
OECD / Eurostat / WHO	Spesa sanitaria privata (spesa diretta delle famiglie e schemi volontari di finanziamento) stimata secondo il System of Health Accounts (SHA 2011)	ICHA	45,9	1970-2023	Sì	Internazionale e nazionale
MEF	Spesa sanitaria privata sostenuta dai cittadini e trasmessa al Sistema Tessera sanitaria (HCS)	Spesa per regione, mittente e tipologia	43,1	2016-2023	No	Nazionale e regionale

Fonte: elaborazione dell'Autrice su dati pubblicati in Del Vecchio *et al*. (2024)

regionale è cruciale per studiare differenze territoriali e disparità tra sistemi sanitari regionali, ma un elemento critico è che solo due di queste fonti consentono stime a questo livello: l'Indagine sulle spese delle famiglie (Istat)[3], uno strumento importante per documentare la spesa sanitaria delle famiglie, che tuttavia risente di variabilità campionaria e di possibili fenomeni di sotto-dichiarazione; e, sul versante amministrativo, la base dati del Sistema Tessera sanitaria[4], oggi la fonte più ampia disponibile con dettaglio regionale, che registra le transazioni sanitarie trasmesse elettronicamente dai fornitori. Tuttavia, nelle statistiche pubblicate, il potenziale informativo del STS rimane fortemente compresso: la scomposizione per categorie di spesa è limitata e mancano elementi che permettano di distinguere i pagamenti OOP dai rimborsi mediati da assicurazioni, casse mutualistiche o altri schemi. Questa distinzione è cruciale perché i diversi meccanismi di finanziamento privato non sono equivalenti né sul piano economico né su quello distributivo. I pagamenti OOP possono tradursi direttamente in barriere economiche e, per alcuni gruppi, in rinuncia o ritardo nelle cure. Il finanziamento assicurativo, invece, può ridurre l'esposizione al momento dell'utilizzo, ma può anche amplificare le disuguaglianze, concentrando i benefici tra chi può sostenere economicamente il pagamento dei premi e contribuendo a segmentare l'accesso ai servizi. Evidenze recenti per l'Italia mostrano inoltre che la relazione tra copertura assicurativa volontaria e spesa sanitaria privata risulta empiricamente complessa da ricostruire e difficilmente osservabile attraverso statistiche aggregate, anche a causa della limitata disponibilità di informazioni granulari che consentano di distinguere tra pagamenti OOP e spesa intermediata da assicurazioni o fondi sanitari (Celidoni e Rebba 2024). In assenza di tale

3 Per informazioni più dettagliate, si veda https://www.istat.it/informazioni-sulla-rilevazione/spese/.

4 Per maggiori informazioni, si veda https://www.rgs.mef.gov.it/VERSIONE-I/attivita_istituzionali/previsione/spesa_sociale/sistema_tessera_sanitaria/index.html.

distinzione, l'analisi empirica resta vincolata e le affermazioni sugli effetti del finanziamento privato risultano difficili da valutare con rigore.

3. Dati sulla copertura assicurativa sanitaria e sulle detrazioni fiscali

La copertura assicurativa sanitaria volontaria presenta una forte eterogeneità tra i paesi europei sia in termini di diffusione sia rispetto alla funzione svolta all'interno dei sistemi sanitari (Sagan and Thomson 2016b; OECD 2024). In alcuni contesti opera come forma sostitutiva di copertura per gruppi esclusi dagli schemi statutari o che possono esercitare un'opzione di uscita dal sistema pubblico; in altri, invece, svolge una funzione complementare/supplementare, offrendo soprattutto un accesso più rapido, una maggiore scelta del fornitore o coperture aggiuntive per servizi non inclusi nel paniere pubblico.

Dal punto di vista delle politiche, l'espansione di questi schemi assicurativi e le strategie di incentivazione al ricorso al settore privato sono spesso giustificate richiamando un effetto di 'decongestione', ossia l'idea che lo spostamento di una quota di domanda verso il privato possa ridurre la pressione sui servizi pubblici e migliorare la capacità di risposta del sistema pubblico per coloro che restano. Tuttavia, valutare se tali effetti si realizzino – e, soprattutto, per quali gruppi e in quali contesti – è empiricamente difficile a causa di problemi di misurazione che riguardano la definizione stessa di copertura assicurativa e la sua comparabilità tra fonti in termini di classificazione delle diverse tipologie di coperture assicurative. In particolare, la tabella 2 mostra che le istituzioni internazionali utilizzano denominazioni in larga misura simili, ma adottano tassonomie differenti, generando incoerenze nel modo in cui gli schemi assicurativi sono codificati nei dataset e, di conseguenza, nei tassi di copertura riportati. La reportistica OCSE distingue tra assicurazione sanitaria 'primaria', 'duplicativa', 'complementare' e 'supplementare', mentre la tassonomia comunemente impiegata nei rapporti dell'European Observatory raggruppa gli schemi in categorie 'sostitutiva', 'complementare' e 'supplementare', con confini in parte differenti (Sagan e Thomson 2016a). Differenze classificatorie possono tradursi in discrepanze anche ampie nei livelli di copertura riportati e, quindi, incidere non solo sulla comparabilità internazionale, ma anche sull'interpretazione del ruolo svolto dall'assicurazione nei sistemi sanitari nazionali[5].

In Italia, la sfida di misurazione è aggravata dalla limitata disponibilità di microdati. Esistono strumenti di indagine – in particolare, l'Indagine Multiscopo Istat sulle famiglie *Aspetti della vita quotidiana,* una rilevazione campionaria annuale strutturata come *repeated cross-sectio*n – che includono quesiti sul possesso di una polizza sanitaria da parte dell'intervistato o di altri componenti del nucleo familiare. Tuttavia, tale informazione, pur presente

Tabella 2. Confronto tra tassonomie della VHI (OCSE vs. European Observatory)

Tassonomia OCSE	Tassonomia European Observatory
VHI primaria: unico canale di accesso alla copertura sanitaria perché non esiste copertura pubblica/ obbligatoria oppure gli individui non sono eleggibili oppure possono optare per l'uscita dallo schema pubblico.	*VHI sostitutiva*: copre persone escluse dalla copertura finanziata pubblicamente o autorizzate a optare per l'uscita.
VHI duplicativa: copre servizi già inclusi nella copertura pubblica/obbligatoria offrendo accesso a fornitori diversi o a livelli di servizio più elevati (ad es. accesso più rapido), senza esentare dagli obblighi contributivi.	*VHI supplementare*: offre accesso più rapido ai servizi, maggiore scelta del fornitore o migliori comfort/amenità.
VHI complementare: copre in tutto o in parte i costi residui non altrimenti rimborsati (ad es. compartecipazioni, ticket). *VHI supplementare:* fornisce copertura per servizi sanitari aggiuntivi non coperti dallo schema pubblico/ obbligatorio.	*VHI complementare*: copre servizi esclusi dal paniere di prestazioni finanziate pubblicamente e/o i costi a carico dell'utente per beni e servizi inclusi nel paniere pubblico.

Fonte: elaborazione dell'Autrice su dati OECD (2025) e Sagan e Thomson (2016a)

5 Un esempio è il Belgio: nel 2016 l'OCSE indicava che gli schemi di VHI complementare e supplementare coprivano, rispettivamente, il 97,9% e il 48,6% della popolazione. Sagan e Thomson (2016b), adottando la loro tassonomia, stimavano invece una copertura supplementare pari a circa il 60%.

nel questionario, non è inclusa nei microdati ad uso pubblico rilasciati da Istat[6]. Ciò preclude analisi distributive centrali per una valutazione orientata all'equità, quali l'identificazione dei gruppi socioeconomici e dei territori più esposti alla copertura e la verifica se l'assicurazione sanitaria sia associata a specifici profili di spesa privata e di accesso all'offerta privata.

Gli incentivi fiscali rappresentano un'ulteriore dimensione strettamente connessa del finanziamento privato. A livello internazionale, banche dati quali MISSOC e PwC Worldwide Tax Summaries documentano una marcata eterogeneità nelle agevolazioni fiscali relative ai premi assicurativi sanitari e alla spesa sanitaria[7]. L'Italia è un caso particolarmente istruttivo. Sul versante individuale, lo strumento principale è una detrazione IRPEF del 19% sulle spese sanitarie ammissibili oltre una franchigia (129,11 euro). Sul versante datoriale, i contributi ai fondi sanitari integrativi possono essere esclusi dal reddito imponibile dei lavoratori e interamente deducibili per i datori di lavoro entro un tetto (3.615,20 euro annui), a condizione di registrazione e del rispetto di requisiti di mutualità e solidarietà. Sebbene tali disposizioni possano stimolare copertura e spesa, esse sollevano anche questioni di equità, poiché i vantaggi fiscali tendono a generare benefici maggiori per individui e imprese con basi imponibili più elevate e rapporti di lavoro più stabili[8].

Nonostante la loro rilevanza, i dati sugli incentivi fiscali sono oggi diffusi in forme che ne limitano fortemente l'analisi. Il Ministero dell'Economia e delle finanze rende disponibili soltanto aggregati regionali relativi alla spesa sanitaria detraibile al 19%, ma tali valori non distinguono tra ticket del SSN e spese sostenute interamente nel settore privato. Questa mancanza di disaggregazione informativa impedisce di valutare in modo convincente la natura redistributiva dello strumento: se le detrazioni finanzino principalmente forme di compartecipazione al sistema pubblico, se finiscano per incentivare l'utilizzo di prestazioni private, oppure se avvantaggino in misura sproporzionata famiglie già più avvantaggiate.

Tali carenze informative sono strettamente intrecciate con le altre criticità discusse. In assenza di dati individuali capaci di separare ticket del SSN, spese per prestazioni integralmente private e rimborsi, non è possibile stimare con rigore l'impatto degli incentivi fiscali né analizzare in che misura la copertura assicurativa si associ ai profili di spesa privata. In questo quadro, il Sistema Tessera Sanitaria rappresenta un'infrastruttura amministrativa potenzialmente decisiva, poiché raccoglie informazioni sulle transazioni sanitarie trasmesse elettronicamente dai fornitori a fini fiscali e di rendicontazione. In linea di principio, i registri STS potrebbero sostenere una classificazione più coerente delle spese per tipologia di servizio e contesto di erogazione, consentire di distinguere i ticket dalle prestazioni acquistate privatamente e, laddove l'informazione fosse disponibile, identificare se e attraverso quali canali le spese siano rimborsate (assicurazioni, fondi/casse mutualistiche o altri schemi). In particolare, l'accesso ai microdati STS permetterebbe analisi distributive oggi fuori portata: l'incidenza dei benefici fiscali per gruppi di reddito e territori, l'interazione tra detrazioni e copertura assicurativa, e il grado in cui il finanziamento privato attenui o amplifichi le disuguaglianze di accesso. Al momento, tuttavia, questo potenziale risulta solo parzialmente valorizzato: pur offrendo totali utili, le elaborazioni STS oggi pubblicate restano aggregate e non rendono disponibili dati a livello individuale che consentano di ricostruire e separare ticket del SSN, spese per prestazioni integralmente private e rimborsi. Ne consegue che molte domande di policy, centrali nel dibattito, rimangono empiricamente poco esplorate.

Conclusioni e implicazioni di policy

Questo lavoro descrive il panorama informativo italiano su spesa sanitaria privata, assicurazione

6 Un quesito analogo è presente anche nell'indagine Istat *Condizioni di salute e ricorso ai servizi sanitari* (modulo italiano dell'European Health Interview Survey (EHIS)) disponibile per il 2019. Tale informazione è stata utilizzata, ad esempio, da Brenna (2025); tuttavia, la sua disponibilità, attualmente limitata a un singolo anno, non consente analisi longitudinali e ne riduce l'utilità empirica per valutare dinamiche temporali della copertura assicurativa.

7 Per informazioni più dettagliate, si veda il database MISSOC e PwC Worldwide Tax Summaries.

8 Diversi studi empirici offrono indicazioni sugli effetti distributivi delle detrazioni fiscali per spese sanitarie in Italia. Ad esempio, Di Novi *et al.* (2018) utilizzano un modello di microsimulazione del sistema di imposte e trasferimenti per valutare l'impatto delle detrazioni sanitarie, mostrando che l'attuale configurazione tende a favorire i segmenti più abbienti della popolazione e risulta relativamente meno efficace nel sostenere i soggetti a basso reddito. Analogamente, Brenna (2018), basandosi su dati aggregati a livello regionale, conclude che i benefici delle detrazioni si concentrano in misura sproporzionata tra contribuenti con redditi più elevati e livelli di istruzione maggiori. Infine, Cavazza *et al.* (2023), avvalendosi di un dataset originale relativo agli iscritti a un piano di assicurazione sanitaria privata offerto da una grande impresa, evidenziano una marcata eterogeneità territoriale nell'adesione, con una copertura più concentrata nelle regioni settentrionali e relativamente più ricche.

sanitaria volontaria e incentivi fiscali legati alla spesa sanitaria come tre dimensioni interconnesse del finanziamento privato che incidono in modo crescente sui dibattiti relativi ad accesso, capacità di risposta ed equità. L'analisi mostra che la base informativa italiana resta frammentata e che tale frammentazione non è un mero problema tecnico, ma un vincolo sostanziale alla valutazione empirica e, quindi, alla qualità della discussione pubblica e della decisione di policy. I livelli di spesa sanitaria privata variano tra le fonti istituzionali perché poggiano su framework contabili, perimetri di riferimento e scelte di confine differenti; inoltre, le statistiche disponibili raramente offrono stime regionali che scompongano la spesa out-of-pocket da quella riconducibile all'assicurazione sanitaria, e mancano dati individuali che consentano la stessa scomposizione in modo sistematico. Ne deriva che fenomeni potenzialmente diversi – pagamenti diretti per prestazioni fuori dal perimetro pubblico, compartecipazioni al SSN, spesa mediata o rimborsata da schemi assicurativi – tendono a essere osservati in modo aggregato o indistinto, riducendo la capacità di interpretare correttamente l'evoluzione del finanziamento privato e i suoi effetti sulla protezione finanziaria delle famiglie.

Anche la copertura assicurativa privata è difficile da confrontare a livello internazionale e, in Italia, le sue implicazioni distributive sono complesse da valutare per via dell'accesso limitato ai microdati pertinenti. Tale lacuna pesa in modo particolare perché l'assicurazione sanitaria privata rappresenta un meccanismo che può interagire con l'offerta pubblica e con i comportamenti di ricorso al privato, amplificando o attenuando disuguaglianze preesistenti. Senza informazioni granulari, non è possibile stabilire in modo credibile se e quanto l'assicurazione si concentri in specifici gruppi socioeconomici o territori, né valutare se essa operi prevalentemente come integrazione al settore pubblico o come canale di accesso alternativo che può segmentare ulteriormente l'utilizzo dei servizi.

Infine, le evidenze sulle detrazioni fiscali sono in larga parte aggregate e non distinguono i ticket del SSN dalla spesa pienamente privata; ciò limita la possibilità di valutare con precisione l'incidenza degli incentivi e di comprendere se il disegno attuale produca effetti regressivi, premiando soprattutto chi ha maggiore capacità contributiva e maggiore propensione/possibilità di spesa. Nel complesso, questi limiti hanno implicazioni pratiche perché restringono ciò che può essere affermato con credibilità su chi paga, chi beneficia e su come il finanziamento privato interagisca con le disuguaglianze tra gruppi di reddito e territori.

Pertanto, il rafforzamento della misurazione e dell'accesso a dati più granulari sono prerequisiti fondamentali per un monitoraggio significativo e per la valutazione delle politiche in sistemi sanitari sempre più ibridi. Questo contributo individua tre priorità tra loro strettamente connesse per consolidare la base informativa e orientare in modo credibile la progettazione delle politiche. Innanzitutto, una misurazione più informativa della spesa sanitaria privata richiede una disaggregazione più fine – almeno a livello regionale e, idealmente, individuale – insieme a una scomposizione trasparente della spesa privata in pagamenti OOP, compartecipazioni e spesa riconducibile all'assicurazione sanitaria privata. Questo passaggio è importante non solo per migliorare la comparabilità tra fonti, ma soprattutto per rendere osservabili i meccanismi: ad esempio, distinguere se una crescita della spesa privata sia trainata da quote elevate di ticket, da farmaceutica, da diagnostica, o da prestazioni private per tempi di attesa.

Inoltre, un migliore accesso ai microdati sulla copertura assicurativa è essenziale per consentire analisi distributive robuste su chi è coperto, dove la copertura si concentra e come essa si correli ai profili di spesa privata e di accesso: quesiti che non possono essere affrontati con gli aggregati attualmente disponibili e che sono centrali per valutazioni di equità e di policy.

Infine, l'accesso ai microdati sulle detrazioni fiscali per spese sanitarie è necessario per valutare adeguatamente l'incidenza degli incentivi e la loro interazione sia con la spesa OOP sia con gli schemi assicurativi privati: solo così è possibile capire se gli incentivi operino come strumento di protezione finanziaria o, viceversa, come sussidio implicito a consumi privati concentrati tra gruppi avvantaggiati.

In questa prospettiva, il STS italiano si configura come un'infrastruttura potenzialmente cruciale. Esso, infatti, registra informazioni dettagliate sulle transazioni sanitarie e potrebbe quindi supportare una scomposizione molto più coerente della spesa privata, abilitando al contempo analisi regionali e individuali, sotto una governance sicura e stringenti tutele della privacy. Se opportunamente governato, l'accesso ai microdati contenuti in tale sistema permetterebbe di rispondere a domande che oggi restano fuori portata: quantificare la quota di spesa privata effettivamente 'evitabile' se l'accesso pubblico fosse più tempestivo; misurare la concentrazione della spesa per fasce di reddito e profili demografici; distinguere con precisione la spesa netta delle famiglie al netto dei rimborsi assicurativi e dei benefici fiscali; verificare la sovrapposizione tra copertura assicurativa e fruizione di prestazioni private.

Sbloccare questo potenziale rafforzerebbe,

dunque, in modo sostanziale la capacità di monitorare il finanziamento privato e di valutare gli incentivi fiscali e le dinamiche di copertura in un SSN in cui componenti pubbliche e private risultano sempre più intrecciate. In termini più generali, migliorare la misurazione non è un obiettivo fine a sé stesso ma è condizione necessaria per sottoporre a verifica empirica affermazioni molto diffuse nel dibattito pubblico odierno che spesso non sono suffragate da evidenze robuste. Infatti, una infrastruttura informativa più solida consentirebbe di stimare con maggiore robustezza gli effetti distributivi, di identificare aree e gruppi più esposti a oneri finanziari e barriere di accesso, e di orientare in modo più mirato gli strumenti di policy. In un sistema sanitario universalistico, questa capacità valutativa è parte integrante della governance: rende possibile un controllo più informato sulle scelte di finanziamento, chiarisce *trade-off* e priorità e riduce il rischio che l'espansione di canali privati avvenga senza strumenti adeguati a monitorarne effetti e conseguenze.

Bibliografia

Besley T., Hall J., Preston I. (1999), The demand for private health insurance: do waiting lists matter?, *Journal of Public Economics*, 72, n.2, pp.155-181

Brenna, E. (2018). Rich and well educated: are these requirements necessary to claim healthcare tax credits in Italy?, *Applied Health Economics and Health Policy*, 16, n.2, pp.207-217

Brenna E. (2025), Public versus private access in the Italian NHS-The use of propensity score matching to provide more insight on the increasing adoption of voluntary health insurance, *Health Policy*, 154 <https://doi.org/10.1016/j.healthpol.2025.105271>

Cavazza M., Del Vecchio M., Fattore G., Fenech L. (2023), Geographical variation in the use of private health insurance in a predominantly publicly funded system, *Health Policy*, 130

<DOI: 10.1016/j.healthpol.2023.104720>

Celidoni M., Rebba V. (2024), Spesa sanitaria privata e assicurazioni integrative: alcune evidenze dall'indagine SHARE, *Rivista di Politica Economica*, 103, n.1, pp.203-241

Del Vecchio M., Fenech L., Preti L. M., Rappini V. (2024), I consumi privati in sanità, in CERGAS - Bocconi (a cura di), *Rapporto OASI 2024,* Milano, Egea

Di Novi C., Marenzi A., Rizzi D. (2018), Do healthcare tax credits help poor-health individuals on low incomes?, *The European Journal of Health Economics*, 19, n.2, pp.293-307

OECD (2024), *Health at a Glance: Europe 2024: State of Health in the EU Cycle*, Paris, OECD Publishing <https://doi.org/10.1787/b3704e14-en>

OECD (2025), *Voluntary Health Insurance. OECD Health Statistics 2025: Definitions, Sources and Methods,* Paris, OECD Publishing

Propper C. (2000), The demand for private health care in the UK, *Journal of health economics*, 19, n.6, pp.855-876

Propper C., Green K. (2001), A larger role for the private sector in financing UK health care: the arguments and the evidence, *Journal of Social Policy*, 30*,* n.4, pp.685-704

Purcel A. A., Dragos C. M., Mare C., Dragos S. L. (2023), Voluntary health insurance and out-of-pocket payments in European OECD countries, *Economic Modelling,* 120, pp.106-190

Sagan A., Thomson S. (2016a), *Voluntary Health insurance in Europe: Country Experience,* Copenhagen, European Observatory on Health Systems and Policies

Sagan A., Thomson S. (2016b), *Voluntary Health Insurance in Europe: Role and Regulation,* Copenhagen, European Observatory on Health Systems and Policies

World Health Organization (2023), *Can people afford to pay for health care? Evidence on financial protection in 40 countries in Europe: summary,* Copenhagen, WHO Regional Office for Europe

Clara Andresciani
clara.andresciani@uniroma1.it
Dottoressa di ricerca in Economia politica presso Sapienza Università di Roma. La sua ricerca si colloca tra economia sanitaria ed economia delle diseguaglianze, con un'attenzione particolare alle trasformazioni del rapporto tra settore pubblico e settore privato nei sistemi sanitari. Nel corso del dottorato ha studiato come la configurazione di questo mix influenzi l'accesso alle cure e, in particolare, l'emergere di bisogni sanitari insoddisfatti, sia a livello europeo sia nel contesto italiano.

La profilazione delle persone disponibili al lavoro in Piemonte

Il ruolo dei dati amministrativi e dei dati territoriali locali

Monica Bosia
Agenzia Piemonte Lavoro

Marianna Carlone
Università degli Studi di Torino

Elisa Floredan
Agenzia Piemonte Lavoro

Francesca Prever
Università degli Studi di Torino

Il lavoro analizza il potenziale dell'integrazione tra dati amministrativi individuali e indicatori territoriali per il miglioramento dei modelli di profiling dei Servizi per l'impiego. Utilizzando dati delle schede anagrafico-professionale piemontesi e un modello logit di riferimento Anpal, si mostra che l'introduzione di variabili dei Sistemi locali del lavoro e l'utilizzo di tecniche di machine learning migliorano la capacità predittiva e la sensibilità nella previsione del mancato reinserimento occupazionale, con implicazioni per le politiche attive del lavoro.

Profiling people available for work in Piedmont. The role of administrative data and local territorial data

This paper analyses the integration of individual administrative data and local labour market indicators to improve profiling models used in employment services. Using Piedmont SAP data and a benchmark Anpal logit model, we show that including Local Labour Systems variables and machine learning techniques enhances predictive performance and sensitivity in identifying individuals at risk of non-employment, with relevant implications for active labour market policies.

DOI: 10.53223/Sinappsi_2026-02-10

Citazione
Bosia M., Carlone M., Floredan E., Prever F. (2026), La profilazione delle persone disponibili al lavoro in Piemonte. Il ruolo dei dati amministrativi e dei dati territoriali locali, *Sinappsi*, XVI, n.2, pp.145-160

Parole chiave
Dati amministrativi
Mercato del lavoro
Occupabilità

Keywords
Administrative data
Labour market
Employability

Introduzione

Il progressivo spostamento dell'analisi economica dai dati campionari ai dati amministrativi rappresenta uno dei cambiamenti strutturali più rilevanti nella ricerca empirica contemporanea. Nelle principali riviste internazionali – quali *American Economic Review* e *Quarterly Journal of Economics* – la quota di articoli basati su microdati d'indagine si è drasticamente ridotta tra il 1980 e il 2010, mentre è cresciuta in misura speculare la produzione fondata su archivi amministrativi (Chetty 2012). Anche nel contesto italiano tale dinamica è evidente: una larga parte degli studi empirici sul mercato del lavoro utilizza oggi basi dati amministrative, in particolare quelle dell'Inps, a fronte di un ricorso relativamente più contenuto all'Indagine sulle Forze di lavoro dell'Istat (Leombruni 2024).

Questo mutamento non è soltanto quantitativo, ma riflette una trasformazione più profonda nelle esigenze conoscitive della ricerca economica e sociale. La crescente attenzione verso l'eterogeneità individuale, le micro-dinamiche e gli effetti distributivi delle politiche – in particolare nell'ambito delle politiche del lavoro e del welfare – alimenta una domanda strutturale di microdati. La valutazione causale degli interventi pubblici richiede infatti informazioni granulari sui percorsi individuali, non sempre osservabili attraverso indagini campionarie. In questa prospettiva, i microdati assumono le caratteristiche di un bene pubblico, il cui valore scientifico dipende anche dalla possibilità di accesso e replicabilità delle analisi (Trivellato 2019).

Il tema dell'accesso e dell'integrazione delle fonti informative si colloca, tuttavia, all'incrocio tra esigenze di ricerca e tutela della riservatezza.

La possibilità di collegare archivi amministrativi a popolazioni statisticamente ben definite costituirebbe una base solida per l'inferenza, ma – anche alla luce delle normative sulla protezione dei dati – tale ancoraggio non è sempre disponibile ai ricercatori. Nel contesto italiano, l'utilizzo degli archivi amministrativi per la ricerca ha radici consolidate ma ha conosciuto una nuova accelerazione grazie, da un lato, all'introduzione delle Comunicazioni Obbligatorie (COB) che tracciano in modo sistematico avvii, trasformazioni e cessazioni dei rapporti di lavoro e, dall'altro, alla crescente attenzione per le disuguaglianze e i determinanti sociali delle condizioni economiche e occupazionali (Leombruni e Ardito 2026). Tutto ciò ha stimolato diverse iniziative volte a promuovere l'uso dei dati amministrativi su occupazione, welfare, istruzione e salute a fini di ricerca.

Il database COB contiene infatti informazioni estremamente ricche, con variabili relative al lavoratore – quali dati demografici di base e livello di istruzione – e informazioni molto dettagliate sul rapporto di lavoro, tra cui la tipologia contrattuale, il contratto collettivo applicato, una classificazione occupazionale molto fine e il settore di attività. La raccolta e la gestione di tali comunicazioni avvengono attraverso sistemi informativi regionali o tramite il Sistema nazionale. In Piemonte, ad esempio, le COB sono gestite attraverso la piattaforma GE.CO (Gestione online Comunicazioni Obbligatorie), mentre altre regioni utilizzano direttamente il portale nazionale del Ministero del Lavoro. La disponibilità di informazioni tempestive e riferite all'intera popolazione ha favorito una rapida espansione degli studi sul mercato del lavoro italiano, in particolare a livello regionale. Molte regioni italiane consentono oggi l'accesso scientifico ai propri archivi COB regionali, rendendo possibile la produzione di analisi dettagliate e aggiornate sulle dinamiche dei mercati del lavoro locali.

In tale scenario si inserisce il presente lavoro, che esplora l'impiego integrato di dati amministrativi provenienti da Agenzia Piemonte Lavoro (APL) e di informazioni territoriali dell'Istat a livello di Sistemi locali del lavoro (SLL) per finalità di profiling. La profilazione viene effettuata al momento del rilascio della Dichiarazione di immediata disponibilità (DID), l'atto formale con cui un individuo si dichiara immediatamente disponibile al lavoro e che costituisce il punto di ingresso nel sistema di osservazione amministrativa utilizzato nello studio. A seguito della DID, i soggetti sono tenuti a sottoscrivere un Patto di servizio personalizzato (PSP) e a partecipare alle politiche attive del lavoro, attraverso la partecipazione a misure definite con il supporto del Centro per l'impiego. Tali interventi comprendono un insieme di servizi (LEP – Livelli essenziali delle prestazioni) calibrati in funzione del livello di occupabilità e di fragilità risultante dalla profilazione quantitativa e dall'assessment qualitativo (Anpal[1] 2022).

L'obiettivo è verificare se l'utilizzo di flussi amministrativi ancorati a contesti locali consenta di migliorare la capacità predittiva e discriminante dei modelli di profiling rivolti ai soggetti in cerca di occupazione oggi utilizzati a livello nazionale. L'ipotesi di fondo è che una maggiore aderenza al contesto territoriale effettivo in cui si svolgono le transizioni lavorative possa cogliere meglio eterogeneità e dinamiche locali, rafforzando la qualità informativa della profilazione e, conseguentemente, il potenziale supporto alle politiche attive del lavoro.

Il presente articolo si sviluppa secondo il seguente schema. La sezione 1 illustra il ruolo e le caratteristiche dei dati amministrativi, con particolare attenzione al quadro normativo italiano. La sezione 2 ricostruisce l'evoluzione del profiling nei Servizi per l'impiego, a livello internazionale e nazionale, e descrive il modello Anpal. La sezione 3 espone la strategia empirica e confronta il modello nazionale con la replica su dati regionali piemontesi, attraverso l'analisi degli effetti marginali su profili rappresentativi. La sezione 4 estende l'analisi introducendo indicatori di contesto territoriale a livello di Sistema locale del lavoro, valutandone il contributo alla capacità predittiva. La sezione 5 approfondisce l'impiego di tecniche di machine learning (Random Forest e Boosting). L'ultima sezione conclude discutendo le implicazioni di policy e le prospettive di sviluppo.

1. I dati amministrativi come risorsa per la ricerca e le politiche pubbliche

Caratteristiche, potenzialità e quadro normativo

I dati di natura amministrativa costituiscono un capitale informativo strategico per la progettazione, l'attuazione, il monitoraggio e la valutazione delle politiche pubbliche, in quanto sono caratterizzati da elevata copertura informativa, granularità dei dati e aggiornamento continuo. Le principali fonti comprendono i registri civili, utilizzati per le analisi

1 Agenzia nazionale per le Politiche attive del lavoro (Anpal) (2022), *Delibera n. 5 del 9 maggio 2022, Approvazione strumenti per la profilazione quantitativa e qualitativa nell'attuazione dell'assessment, standard dei servizi e costi standard, Banca Dati del Lavoro. https://bdl.sviluppolavoroitalia.it/bdds/ViewScheda.action?product=NORMALAVORO&uid=b3b7658c-3c87-44c9-b85b-ea1eef776f33&title=scheda.*

demografiche e della struttura della popolazione, i dati fiscali, impiegati per lo studio delle dinamiche economiche, e, soprattutto per quanto qui rileva, i registri sull'impiego, che consentono di analizzare il funzionamento del mercato del lavoro e di valutare l'efficacia delle politiche occupazionali. L'impulso del Legislatore è stato determinante nella costruzione di questo patrimonio informativo, inizialmente finalizzato a migliorare l'accesso ai servizi pubblici e a semplificare i procedimenti amministrativi attraverso l'uso delle tecnologie digitali. Un passaggio fondamentale è rappresentato dalla legge n. 59/1997[2] (c.d. legge Bassanini), cui ha fatto seguito l'adozione del Codice dell'amministrazione digitale (CAD) nel 2005[3], che ha definito il quadro giuridico per la digitalizzazione della pubblica amministrazione e per la condivisione dei dati tra enti.

Il Codice ha inoltre introdotto infrastrutture e principi volti a favorire l'interoperabilità dei sistemi informativi pubblici, tra cui la Piattaforma digitale nazionale dati (art. 50-ter CAD) e i principi del *digital first* e del *once only*[4]. In questo contesto si colloca anche il ruolo dell'Agenzia per l'Italia digitale[5], responsabile della definizione delle linee guida tecniche per la gestione, la conservazione e l'accesso ai dati della Pubblica amministrazione, secondo standard che ne garantiscano integrità, sicurezza e tracciabilità.

In linea con i principi dell'open government, orientati a garantire trasparenza e integrità nei processi decisionali pubblici[6], le amministrazioni sono tenute a rendere disponibili in formato open data le informazioni di interesse pubblico prodotte nell'ambito dell'attività amministrativa. Questo quadro normativo, rafforzando l'accountability delle istituzioni, favorisce anche l'accesso ai dati per finalità di ricerca. Ulteriori disposizioni regolano l'utilizzo dei dati amministrativi a fini statistici e scientifici: il D.Lgs. n. 322/1989[7] prevede l'uso prioritario degli archivi amministrativi per statistiche ufficiali, mentre il Regolamento (UE) 2016/679[8] e la Direttiva Comstat n. 11/2018 disciplinano la tutela dei dati personali e le modalità di accesso ai microdati anonimizzati per scopi di ricerca.

Dati amministrativi: qualità, limiti e valore strategico per le politiche pubbliche

Pur essendo raccolti nell'ambito dell'attività amministrativa, tali dati possono essere efficacemente utilizzati anche per finalità statistiche e di ricerca, consentendo analisi basate su archivi informativi già esistenti. La registrazione sistematica delle informazioni sull'intero universo di riferimento consente di disporre di basi dati molto estese, che rendono possibili analisi dettagliate su popolazioni e territori e riducono la dipendenza da rilevazioni campionarie, pur richiedendo specifiche attenzioni in termini di qualità e copertura delle informazioni. La qualità dei dati può infatti variare in funzione delle modalità di raccolta e aggiornamento, e la presenza di errori di imputazione, informazioni mancanti o non aggiornate può influenzare l'affidabilità delle analisi statistiche e delle evidenze empiriche. Per questo motivo assumono particolare rilevanza i protocolli di verifica e di valutazione della qualità come quelli adottati nell'ambito del Sistema Statistico Nazionale e dell'Istat, che prevedono criteri standardizzati per la misurazione di dimensioni quali rilevanza, accuratezza, completezza, tempestività, coerenza e accessibilità dei dati[9].

Nonostante l'ampiezza e la granularità, i dati amministrativi presentano limiti nella copertura informativa: non intercettano fenomeni sociali per i quali non è necessario né obbligatorio rivolgersi al pubblico servizio. Per questo, è spesso opportuno integrarli con fonti campionarie o qualitative, adottando disegni di ricerca misti che consentano una comprensione più ampia di fenomeni complessi (Istat 2014).

In Piemonte come in altre regioni d'Italia (Bazzoli *et al.* 2018) l'integrazione di fonti amministrative e campionarie è stata utilizzata per leggere le trasformazioni del mercato del lavoro a livello territoriale, analizzare i percorsi individuali di accesso alle politiche attive, i fabbisogni di competenze espressi dal tessuto produttivo, e orientare di conseguenza la programmazione di interventi formativi e di accompagnamento al lavoro più mirati. In questo senso, l'analisi dei dati amministrativi si configura come strumento di raccordo tra sistema

2 L. 15 marzo 1997 n. 59.

3 D.Lgs. 7 marzo 2005 n. 82 (e successive modificazioni).

4 Il principio *digital first* implica che i servizi pubblici siano erogati prioritariamente in modalità digitale; il principio *once only* stabilisce che le informazioni personali debbano essere comunicate una sola volta allo Stato.

5 Per le linee guida sulla conservazione, si veda: Agenzia per l'Italia digitale (2021), disponibile su: www.agid.gov.it/piattaforme/conservazione.

6 Cfr. Dipartimento della Funzione Pubblica (2024), disponibile su: open.gov.it.

7 D.Lgs. 6 settembre 1989 n. 322.

8 Regolamento (UE) 2016/679 del Parlamento europeo e del Consiglio 27 aprile 2016.

9 Sistan, La qualità dei dati amministrativi per la statistica, https://www.sistan.it/index.php?id=424.

del lavoro, formazione e servizi alle persone e alle imprese. Gli archivi amministrativi non sono solo un supporto digitale, ma una risorsa strategica: la loro interpretazione amplifica le opportunità di incontro tra domanda e offerta di lavoro e abilita modelli decisionali *evidence based*, rafforzando sia il monitoraggio e la valutazione delle politiche pubbliche, sia i processi decisionali individuali.

I dati amministrativi impiegati nel modello

I dati amministrativi impiegati nello studio derivano dalla Scheda anagrafico-professionale (SAP), che costituisce la principale unità operativa primaria del sistema informativo regionale[10]. La SAP rappresenta il fascicolo digitale del lavoratore e raccoglie in modo integrato le informazioni anagrafiche, la biografia professionale completa, il percorso formativo e la storia occupazionale del cittadino, inclusa l'eventuale partecipazione a misure di politica attiva. L'archivio della SAP è articolato in diverse sezioni tra loro collegate: dati anagrafici e di contatto, titoli di studio e formazione, storia lavorativa, stato occupazionale e amministrativo e sezione dedicata alle politiche del lavoro. Quest'ultima comprende, tra le altre informazioni, lo stato di disoccupazione, la Dichiarazione di immediata disponibilità (DID), il Patto di servizio personalizzato (PSP), le azioni concordate con il Centro per l'impiego nell'ambito delle politiche attive del lavoro (ad esempio percorsi collegati al Programma GOL o alla NASpI) e gli esiti della profilazione del lavoratore.

La SAP è alimentata automaticamente dalle Comunicazioni Obbligatorie (COB) e dalle attività dei Centri per l'impiego per quanto concerne le politiche attive del lavoro. I dati vengono aggiornati automaticamente grazie a un sistema di cooperazioni applicative tra archivi regionali e nazionali: il sistema informatico regionale intercetta il codice fiscale del lavoratore e aggiorna automaticamente le informazioni presenti nell'archivio. Gli operatori dei Centri per l'impiego accedono a tali informazioni attraverso il Sistema informativo lavoro Piemonte (SilP).

Profiling – literature review

Negli ultimi decenni, i Servizi pubblici per l'impiego (SPI) hanno assunto un ruolo sempre più centrale nell'implementazione delle politiche attive del lavoro, in un contesto caratterizzato da risorse pubbliche limitate, elevata eterogeneità dei disoccupati e crescenti aspettative in termini di efficacia degli interventi. In questo scenario, il profiling si è affermato come uno strumento chiave di supporto decisionale, volto a migliorare il targeting delle misure di attivazione e a orientare l'allocazione dei servizi in funzione delle caratteristiche individuali dei beneficiari. In senso generale, il profiling consiste nel processo di raccolta e utilizzo sistematico di informazioni sui disoccupati per valutarne le prospettive di reinserimento lavorativo e definire, di conseguenza, l'intensità e la tipologia degli interventi più appropriati. L'obiettivo principale è duplice: da un lato, individuare tempestivamente le persone a maggior rischio di disoccupazione di lunga durata; dall'altro, evitare un utilizzo inefficiente delle risorse pubbliche su individui che presentano elevate probabilità di reinserimento autonomo nel mercato del lavoro. In tal modo, il profiling rappresenta uno strumento preventivo, volto a ridurre la persistenza della disoccupazione e a migliorare l'efficacia complessiva delle politiche attive (Desiere *et al.* 2019). La letteratura e la pratica internazionale distinguono tradizionalmente tre principali approcci al profiling nei Servizi per l'impiego: il profiling basato su regole amministrative (*rule-based profiling*), quello fondato sul giudizio degli operatori (*caseworker-based profiling*) e il profiling statistico. Il primo utilizza criteri predefiniti, quali età, livello di istruzione o durata della disoccupazione, per segmentare i disoccupati in gruppi di utenza e determinare l'accesso ai servizi. Questo approccio è ampiamente diffuso grazie alla sua semplicità, trasparenza e facilità di implementazione, ma tende a trascurare l'eterogeneità individuale all'interno dei gruppi e può riflettere priorità politiche piuttosto che il rischio effettivo di esclusione dal mercato del lavoro. Il profiling basato sul giudizio del *caseworker* attribuisce invece un ruolo centrale alla valutazione discrezionale degli operatori dei SPI, spesso supportata da strumenti di *assessment* qualitativi o quantitativi. Tale approccio consente di incorporare informazioni difficilmente formalizzabili, come la motivazione o specifiche barriere individuali, ma presenta criticità legate alla soggettività delle valutazioni, alla variabilità tra operatori e ai maggiori costi di implementazione, data la forte dipendenza dalle risorse umane disponibili. Il profiling statistico si propone come un'alternativa più standardizzata, basata sull'utilizzo di modelli quantitativi per stimare la probabilità di esiti occupazionali sfavorevoli, in particolare il rischio di disoccupazione di lunga durata. Rispetto agli approcci tradizionali, esso consente di trattare i beneficiari come individui piuttosto che come membri di gruppi omogenei, favorendo una maggiore personalizzazione dei servizi. Numerosi Paesi hanno progressivamente adottato strumenti di profiling statistico, spesso integrandoli con il giudizio

10 Agenzia Piemonte Lavoro, *Scheda anagrafica e professionale (Sap)*, disponibile su: https://agenziapiemontelavoro.it/scheda-informativa/scheda-anagrafica-e-professionale-sap/.

degli operatori, in un'ottica di complementarità piuttosto che di sostituzione (Loxha e Morgandi 2014).

Le esperienze internazionali mostrano alcune best practice ricorrenti nella progettazione e nell'uso dei sistemi di profiling. In primo luogo, l'efficacia del profiling dipende in misura cruciale dalla qualità e dall'integrazione dei dati amministrativi disponibili, che devono essere aggiornati e sufficientemente ricchi da cogliere le principali dimensioni dello svantaggio occupazionale (Bollens 2022). In secondo luogo, i sistemi più efficaci sono quelli inseriti in quadri istituzionali coerenti, in cui il profiling è parte integrante del processo di presa in carico e di definizione dei percorsi personalizzati. Infine, emerge l'importanza di un utilizzo flessibile dei risultati del profiling, che supporti il lavoro dei *caseworker* senza vincolarlo rigidamente, preservando margini di adattamento alle specificità individuali (Desiere *et al.* 2019). Nella maggior parte dei casi, tali modelli si basano su tecniche statistiche relativamente consolidate e trasparenti, utilizzate per stimare la probabilità di reinserimento lavorativo o, in alternativa, il rischio di disoccupazione di lunga durata entro un determinato orizzonte temporale. Esempi emblematici sono rappresentati dal *Worker Profiling and Reemployment Services* negli Stati Uniti, dal *Job Seeker Classification Instrument* in Australia e dal *Work Profiler* nei Paesi Bassi, tutti accomunati dall'uso di informazioni amministrative individuali relative alla storia lavorativa, al livello di istruzione e alle caratteristiche sociodemografiche. Un'ulteriore best practice riguarda l'aggiornamento periodico dei modelli e la loro integrazione nei sistemi informativi dei SPI. Poiché le relazioni tra caratteristiche individuali ed esiti occupazionali possono variare nel tempo in funzione del ciclo economico e delle trasformazioni del mercato del lavoro, i modelli di profiling statistico vengono regolarmente ricalibrati sulla base dei dati più recenti (O'Connell *et al.* 2009).

Nel contesto italiano, l'adozione di strumenti di profiling si è sviluppata con un certo ritardo rispetto ad altri Paesi europei. Per lungo tempo, il sistema delle politiche attive del lavoro si è basato prevalentemente su criteri amministrativi e su un ruolo limitato dei Servizi per l'impiego. Un primo cambiamento significativo è avvenuto con il programma Garanzia Giovani, che ha introdotto un modello di profilazione volto a classificare i beneficiari in fasce di svantaggio sulla base di caratteristiche individuali, fornendo un supporto strutturato all'orientamento e alla definizione dei percorsi di politica attiva. Un ulteriore passo in avanti è stato compiuto nell'ambito delle riforme avviate con il *Jobs Act*, che ha ridefinito il ruolo dei Servizi per l'impiego e posto l'attivazione al centro delle politiche del lavoro. In questo quadro, la Dichiarazione di immediata disponibilità (DID) e la successiva profilazione dei disoccupati sono diventate passaggi fondamentali per l'accesso alle misure di sostegno e per la stipula dei patti di servizio personalizzati (decreto legislativo n. 150/2015[11]). L'introduzione della DID online a partire dal 2017 ha inoltre contribuito a migliorare significativamente il patrimonio informativo disponibile. Più recentemente, l'Agenzia nazionale per le politiche attive del lavoro[12] (Anpal) ha formalizzato un modello di profilazione quantitativa a supporto dell'attuazione del Programma garanzia occupabilità lavoratori (GOL), basato esclusivamente su variabili individuali osservabili al momento della sottoscrizione della DID, relative alle caratteristiche anagrafiche, formative, lavorative e familiari del disoccupato (Anpal 2022). Questo modello rappresenta oggi il principale riferimento nazionale per la profilazione nei Servizi per l'impiego e costituisce il punto di partenza per ulteriori sviluppi e applicazioni a livello territoriale.

2. Trasferibilità del modello Anpal: un confronto tra scala nazionale e regionale

Dati del modello

Al fine di valutare la trasferibilità del modello di profiling adottato a livello nazionale e di verificare se l'utilizzo di informazioni riferite al contesto locale possa migliorarne la capacità predittiva, è stato stimato un modello logit su dati amministrativi piemontesi analogo a quello proposto da Anpal (2022). Formalmente, la probabilità di mancato reinserimento occupazionale è modellata come:

$$P(Y_i = 1 \mid X_i) = \frac{e^{X_i'\beta}}{1 + e^{X_i'\beta}}$$

dove:

- $Y_i=1$ se l'individuo *i* non risulta occupato nei sei mesi successivi alla sottoscrizione della DID, $Y_i=0$ altrimenti;
- X_i è il vettore delle variabili esplicative osservate al momento della DID;
- β è il vettore dei coefficienti da stimare.

Le variabili esplicative incluse riproducono,

11 D.Lgs. 14 settembre 2015 n. 150.

12 L'Agenzia nazionale per le politiche attive del lavoro (Anpal) è stata soppressa con decorrenza dal 1° marzo 2024. Le relative funzioni sono state attribuite al Ministero del Lavoro e delle politiche sociali, mentre il personale del comparto ricerca è confluito nell'INAPP. Anpal Servizi S.p.A. ha assunto la denominazione di Sviluppo Lavoro Italia S.p.A.

per quanto consentito dai dati disponibili, quelle utilizzate nel modello nazionale di Anpal. Esse comprendono, tra le principali: genere, età, titolo di studio, esperienza lavorativa recente e condizione occupazionale nell'anno precedente.

La tabella A1 in Appendice riporta le stime dei coefficienti relativi alle variabili individuali comuni ai due modelli considerati. Il modello nazionale (Anpal 2022) considera come variabile dipendente il mancato reinserimento a 12 mesi dalla DID e stima un modello logit su un campione di circa 3,4 milioni di DID registrate in Italia nel periodo 2018-2019. Il modello regionale fa invece riferimento al mancato reinserimento a 6 mesi dalla DID, su un campione di 163.433 DID sottoscritte in Piemonte nel 2022. La scelta dell'orizzonte di 6 mesi deriva da un vincolo empirico: l'estensione a 12 mesi avrebbe ridotto in modo significativo la numerosità campionaria, a causa della minore disponibilità di follow-up completi nei dati amministrativi piemontesi, compromettendo la stabilità delle stime e la rappresentatività per alcuni sottogruppi.

La comparabilità tra le due specificazioni è parziale, in quanto alcune covariate presenti nel modello nazionale non sono osservabili nei dati regionali. In particolare, non sono disponibili le informazioni relative alla partecipazione a tirocini e alla qualifica professionale prevalente nei 24 mesi precedenti la DID. Inoltre, il trattamento dell'età differisce tra i due modelli. Nel modello nazionale l'età entra come variabile continua (con trasformazioni funzionali per cogliere non linearità) per l'effetto principale, ed è utilizzata anche in forma categoriale (classi di età) per le interazioni con altre variabili. Nel modello piemontese, invece, l'età è disponibile e viene utilizzata esclusivamente in forma categoriale, sia per l'effetto principale sia per le interazioni. Tali differenze nella disponibilità e nella definizione delle covariate possono influenzare la comparabilità dei coefficienti stimati tra i due modelli. In particolare, l'assenza di alcune informazioni presenti nel modello nazionale può determinare fenomeni di variabile omessa, che suggeriscono cautela nell'interpretazione delle differenze osservate tra i coefficienti.

Successivamente, il modello è stato esteso introducendo variabili di contesto riferite al Sistema locale del lavoro, con l'obiettivo di verificare se l'integrazione tra caratteristiche individuali e informazioni territoriali consenta di migliorare la capacità discriminante del profiling. In una fase ulteriore, l'analisi è stata affiancata dalla stima di modelli predittivi più flessibili, al fine di valutare se approcci di machine learning possano fornire performance superiori rispetto alla specificazione logistica utilizzata nel modello nazionale.

Confronto tra modelli

Poiché i modelli logit sono stimati su popolazioni differenti, la comparabilità diretta dei coefficienti, riportati nella tabella A1 in Appendice, può risultare limitata da differenze nella varianza non osservata. Inoltre, non essendo disponibili i microdati utilizzati per la stima del modello nazionale, non è stato possibile calcolare effetti marginali medi pienamente comparabili tra le due popolazioni. Il confronto deve pertanto essere interpretato in termini descrittivi e relativi, privilegiando il segno, la significatività e i pattern complessivi degli effetti. Coerentemente con la finalità del lavoro, orientata alla valutazione della capacità di profiling e di identificazione dei soggetti a maggior rischio di mancato reinserimento occupazionale, i risultati non vengono interpretati in termini causali. Per le dimensioni di maggiore interesse sono stati tuttavia calcolati effetti marginali per profili rappresentativi (MER, *Marginal effects at Representative values*), che consentono di valutare come la probabilità predetta di mancato reinserimento vari tra contesti territoriali per tipologie comparabili di individui.

Sulla base di questa impostazione, si confrontano di seguito gli effetti marginali calcolati per profili rappresentativi nei due contesti (tabella 1).

Nel complesso, il modello stimato sui dati piemontesi riproduce in larga parte le regolarità empiriche individuate da Anpal a livello nazionale, ma emergono differenze significative nell'intensità – e talvolta nel segno – degli effetti, che riflettono sia la diversa scala territoriale sia specificità strutturali del mercato del lavoro piemontese.

L'analisi degli effetti marginali con riferimento alla classe under 24 evidenzia un gradiente crescente dell'età per un uomo con le altre caratteristiche fissate al valore di riferimento in entrambi i modelli. Nel modello piemontese la penalizzazione cresce in modo progressivo e relativamente regolare, passando da circa 1 punto percentuale per i 25-29 anni a oltre 12 punti percentuali per gli over 50. Nel modello nazionale, invece, l'effetto è quasi nullo per i 25-29 anni ma aumenta rapidamente a partire dai 30 anni, raggiungendo circa 21 punti percentuali per gli over 50. L'età avanzata rappresenta dunque un fattore di rischio in entrambi i contesti, ma la penalizzazione risulta più intensa e più rapidamente crescente a livello nazionale.

Una dinamica analoga emerge con riferimento al genere. Nel modello piemontese l'effetto diretto dell'essere donna è contenuto (circa +1 punto percentuale in termini di MER), mentre nel modello nazionale risulta sensibilmente più marcato (circa +9 punti percentuali). Considerando la significatività e la struttura dei coefficienti stimati, nel contesto regionale le differenze di genere sembrano

manifestarsi principalmente attraverso interazioni con altre caratteristiche individuali, piuttosto che configurarsi come uno svantaggio generalizzato. Tuttavia, i MER calcolati per una donna di 30-39 anni con figli indicano che, anche tenendo conto dell'interazione tra genere e condizione familiare, in Piemonte l'effetto aggiuntivo della maternità è sostanzialmente nullo (inferiore a un punto percentuale), mentre a livello nazionale raggiunge circa 4-5 punti percentuali. Il vincolo familiare appare dunque più penalizzante nel quadro nazionale, verosimilmente anche in relazione a differenze territoriali nei servizi di conciliazione e nelle strutture del mercato del lavoro, che possono attenuare o amplificare l'impatto della maternità sulle prospettive di reinserimento occupazionale.

Tabella 1. Effetti marginali su profili rappresentativi (MER) per il modello piemontese e nazionale[13]

Variabile/Profilo	Piemonte (pp)	Italia (pp)
Genere e famiglia (30-39 anni)		
Donna	+1,17	+8,67
Figli	-0,39	-1,20
Donna con figli	+0,72	+4,50
Classi di età (rif. Under 24)		
25-29	+1,20	+0,03
30-39	+2,64	+6,10
40-49	+6,31	+14,18
Over 50	+12,18	+20,92
Esperienza lavorativa recente		
40-49 con esperienza	-41,67	-9,20
Over 50 con esperienza	-42,54	-10,32
Durata presenza in Italia (30-39 anni)		
Fino a 12 mesi	-2,59	-0,56
Da 1 a 2 anni	-5,61	-1,52
Oltre 2 anni	-1,40	+1,42
Nato in Italia	+14,02	-2,0

Nota: Effetti marginali calcolati per profili rappresentativi (MER). I valori sono espressi in punti percentuali (pp) e indicano la variazione della probabilità di mancato reinserimento occupazionale associata alla caratteristica considerata; valori positivi corrispondono a una maggiore probabilità di mancato reinserimento. La variabile dipendente identifica la mancata occupazione successiva alla DID (1 = non occupato; 0 = occupato). Nel modello piemontese l'esito è osservato a 6 mesi dalla DID (N = 163433 DID sottoscritte in Piemonte nel 2022), mentre nel modello nazionale Anpal a 12 mesi dalla DID (circa 3,4 milioni di DID sottoscritte in Italia nel biennio 2018-2019).

Fonte: elaborazioni degli Autori; per il modello nazionale, calcoli propri a partire dai coefficienti riportati in Anpal (2022)

13 Gli effetti marginali sono calcolati variando una sola caratteristica rispetto a un profilo di riferimento. Il profilo di riferimento e le variazioni considerate sono specificati per ciascuna sezione della tabella. In tutte le sezioni, le variabili non esplicitamente menzionate sono fissate ai rispettivi valori di riferimento del modello.

- *Genere e famiglia (30–39 anni)*. Profilo di riferimento: uomo di 30-39 anni, cittadino italiano, senza figli, senza esperienza lavorativa recente.
  - Donna: confronto tra donna e uomo (a parità di età, senza figli e senza esperienza lavorativa recente).
  - Figli: confronto tra uomo con figli e uomo senza figli (a parità di età e senza esperienza lavorativa recente).
  - Donna con figli: confronto tra donna con figli e uomo senza figli (a parità di età e senza esperienza lavorativa recente).
- *Classi di età*. Profilo di riferimento: uomo di età inferiore a 25 anni (under 24), cittadino italiano, senza figli, senza esperienza lavorativa recente. Ogni riga mostra l'effetto di appartenere alla classe di età indicata rispetto al profilo di riferimento.
- *Esperienza lavorativa recente*. Profilo di riferimento: uomo di età inferiore a 25 anni (under 24), con esperienza lavorativa nei sei mesi precedenti la DID. Le righe riportano l'effetto dell'appartenenza alla classe di età indicata, in presenza di esperienza lavorativa recente, rispetto al profilo di riferimento.
- *Durata presenza in Italia (30–39 anni)*. Profilo di riferimento: uomo di 30-39 anni, cittadino italiano, senza figli, senza esperienza lavorativa recente. Le righe riportano l'effetto delle diverse durate di presenza in Italia (in mesi/anni) rispetto alla cittadinanza italiana (categoria di riferimento).

Le differenze tra i modelli risultano particolarmente evidenti con riferimento alla durata della presenza in Italia. Sia i coefficienti sia i MER indicano che, rispetto alla cittadinanza italiana, una presenza relativamente recente è associata a una maggiore probabilità di reinserimento occupazionale. Tuttavia, l'intensità dell'effetto è decisamente più marcata in Piemonte. Per un uomo di 30-39 anni, essere presente in Italia da 1 a 2 anni riduce la probabilità di mancato reinserimento di circa 5,6 punti percentuali nel modello regionale, contro 1,5 punti percentuali nel modello nazionale; effetti più contenuti emergono per permanenze inferiori a un anno. Inoltre, mentre nel modello piemontese la categoria "Nato in Italia" è associata a un aumento di circa 14 punti percentuali della probabilità di non reinserimento, nel modello nazionale l'effetto è di segno opposto e di entità modesta. Tale divergenza suggerisce che la variabile catturi nel contesto regionale dinamiche specifiche, plausibilmente legate a meccanismi di selezione migratoria: una quota rilevante di lavoratori stranieri entra infatti nel Paese con un impiego già attivo o con una forte pressione a reinserirsi rapidamente, anche in funzione del rinnovo del permesso di soggiorno.

La differenza più marcata tra i due modelli riguarda tuttavia l'esperienza lavorativa recente. I coefficienti piemontesi risultano molto più elevati in valore assoluto e i MER confermano un impatto estremamente forte: per un uomo di 40-49 anni, aver lavorato nei sei mesi precedenti la DID riduce la probabilità di mancato reinserimento di oltre 40 punti percentuali; un effetto analogo si osserva per gli over 50. Nel modello nazionale, dove l'esperienza è misurata su un orizzonte di dodici mesi, l'effetto si attesta intorno ai 9-10 punti percentuali. La diversa finestra temporale contribuisce certamente a spiegare lo scarto, ma il risultato indica anche che nel contesto piemontese la continuità occupazionale recente rappresenta un segnale particolarmente forte di occupabilità, soprattutto per le classi di età mature.

Per alcune dimensioni, quali il livello di istruzione e il settore di occupazione precedente, il numero elevato di categorie rende meno agevole una sintesi tramite effetti marginali su profili rappresentativi. Per tali dimensioni si fa pertanto riferimento ai coefficienti stimati, riportati nella tabella A1 in Appendice. Per quanto concerne il capitale umano, entrambi i modelli evidenziano che livelli di istruzione più elevati sono generalmente associati a una minore probabilità di disoccupazione di medio periodo. Nel modello piemontese per i titoli universitari, sebbene i coefficienti non siano statisticamente significativi, si osserva un'inversione di segno, a differenza dell'andamento decrescente atteso. Queste evidenze appaiono controintuitive in ragione dei limiti informativi dei dati amministrativi regionali con i quali è stimato il modello. Infatti, nonostante la cooperazione applicativa consenta l'aggiornamento dei dati a livello nazionale e interregionale, ciascuna regione può accedere esclusivamente alle informazioni di propria pertinenza, secondo il principio della minimizzazione dei dati[14]. Ne deriva che un individuo laureato che, successivamente alla DID, trasferisca residenza e domicilio in un'altra regione – come avviene frequentemente nel caso di studenti fuori sede che rientrano nella regione d'origine – e vi trovi occupazione, non risulterà occupato nei dati piemontesi. Questo può determinare una sovrastima della disoccupazione per tale categoria di titolo di studio. Per le categorie con titolo di studio non superiore al diploma, tale meccanismo dovrebbe essere meno rilevante, dal momento che la mobilità interregionale associata ai percorsi formativi interessa soprattutto gli individui che intraprendono studi universitari.

Con riferimento al settore di occupazione precedente, entrambi i modelli mostrano che l'aver maturato un'esperienza lavorativa in specifici comparti è generalmente associato a una minore probabilità di mancato reinserimento rispetto ai soggetti privi di precedenti esperienze lavorative. Nel modello piemontese emerge una maggiore omogeneità nel segno dei coefficienti tra i diversi settori, suggerendo che il semplice mantenimento di un legame recente con il mercato del lavoro costituisca un fattore protettivo rilevante, indipendentemente dal comparto di provenienza. Le interazioni di genere confermano inoltre che alcuni settori presentano dinamiche differenziate tra uomini e donne, con effetti talvolta più accentuati nel contesto regionale.

In sintesi, il confronto tra modello nazionale e modello piemontese indica che le determinanti fondamentali della disoccupazione di medio periodo sono largamente condivise, ma l'intensità degli effetti varia sensibilmente a livello territoriale. Il modello piemontese mette in luce un ruolo particolarmente forte della storia lavorativa recente e di alcune caratteristiche specifiche del contesto locale, mentre nel modello nazionale emergono penalizzazioni più

14 Per la Regione Piemonte i casi accessibili sono:
(i) cittadino residente/domiciliato in Piemonte con COB in azienda localizzata in Piemonte;
(ii) cittadino residente/domiciliato in altra regione con COB in azienda localizzata in Piemonte;
(iii) cittadino residente/domiciliato in Piemonte con COB in azienda localizzata in altra regione.

Tabella 2. Coefficienti stimati per le variabili di contesto selezionate

Variabile	Coefficiente	Significatività
Indipendenti delle imprese individuali per 100 indipendenti	0,0459	***
Indipendenti delle microimprese per 100 indipendenti	-0,1332	***
Dipendenti delle imprese ad alta tecnologia e ad alta intensità di conoscenza per 100 dipendenti	0,0241	***
Indipendenti per 100 residenti (25-34 anni)	0,5093	***
Indipendenti per 100 residenti (totale)	-0,4751	***
Indipendenti 25-34 anni per 100 indipendenti	-0,4547	***
Addetti 25-34 anni per 100 addetti	0,1766	***
Addetti con cittadinanza straniera per 100 addetti (25-64 anni)	0,0192	**
Femmine per 100 indipendenti (25-64 anni, istruzione terziaria)	0,0074	*
Indipendenti per 100 residenti (25-64 anni)	-0,1023	

Nota: Coefficienti stimati mediante modello logit. La specificazione include anche le variabili individuali discusse nel paragrafo 4 (non riportate in tabella). Le variabili di contesto sono state selezionate mediante procedura di best subset selection basata sul criterio AIC. Per la definizione della variabile dipendente e del campione si rimanda alla nota della tabella 1. Gli asterischi indicano il livello di significatività statistica: *** p < 0,01; ** p < 0,05; * p < 0,10.
Fonte: elaborazione degli Autori su dati Agenzia Piemonte Lavoro

marcate legate a età, genere e condizione familiare. Nel complesso, i risultati suggeriscono che le politiche di reinserimento dovrebbero essere non solo personalizzate rispetto al profilo individuale, ma anche calibrate sulle specificità strutturali dei mercati del lavoro regionali.

3. Modello con variabili di contesto

Per approfondire il ruolo delle caratteristiche territoriali nel determinare le probabilità di reinserimento occupazionale, è stato stimato un ulteriore modello logistico che integra, alle variabili individuali già considerate, un insieme di indicatori strutturali riferiti ai Sistemi locali del lavoro (SLL). Gli indicatori territoriali sono tratti dal Sistema informativo sui SLL dell'Istat e derivano dall'integrazione di più registri statistici ufficiali (Istat 2025). In particolare: gli addetti, dipendenti e indipendenti provengono dal Registro statistico delle imprese attive (ASIA); le caratteristiche individuali degli occupati (età, cittadinanza, istruzione) derivano dal Registro integrato degli occupati (RBI); i residenti sono tratti dalla popolazione residente (Anagrafe e bilancio demografico). Gli indicatori utilizzati nel modello (ad esempio, "addetti stranieri per 100 addetti" o "indipendenti per 100 residenti") sono pertanto costruiti da Istat mediante elaborazioni su registri amministrativi integrati[15].

L'obiettivo è verificare se, a parità di caratteristiche sociodemografiche e professionali, il contesto produttivo locale incida in modo significativo sulla probabilità di non rientrare nel mercato del lavoro entro sei mesi dalla sottoscrizione della DID. Mentre il confronto precedente valutava la robustezza dei risultati nel passaggio dalla scala nazionale a quella regionale, questa estensione verifica l'eterogeneità interna al Piemonte. In questa prospettiva, il profiling non può essere inteso come uno strumento neutrale rispetto al territorio, ma deve essere sensibile alle specificità strutturali dei diversi mercati locali, al fine di orientare in modo più mirato le politiche attive.

Per la selezione delle variabili di contesto è stato adottato il metodo della *best subset selection*, utilizzando come criterio di scelta l'*Akaike Information Criterion* (AIC). L'AIC è stato preferito ad altri criteri in quanto, selezionando le specificazioni con migliore capacità predittiva fuori campione, risulta particolarmente adatto per finalità predittive (Akaike 1974; Burnham e Anderson 2002). L'applicazione della procedura ha portato alla selezione di 10 indicatori territoriali, i cui coefficienti stimati sono riportati in tabella 2. Rispetto al modello con sole variabili individuali, il modello arricchito con gli indicatori SLL presenta un AIC sensibilmente più basso, a conferma del miglioramento della capacità predittiva.

L'inclusione di tali variabili evidenzia che il contesto produttivo locale esercita un effetto

15 Per le definizioni di addetto, dipendente e indipendente si rimanda a Istat, Glossario statistico, https://www.istat.it/classificazioni-e-strumenti/glossario/.

autonomo sulla probabilità di non reinserimento. In particolare, una maggiore quota di addetti stranieri è associata a una più elevata probabilità di non reinserimento, il che potrebbe riflettere una maggiore competizione nei segmenti occupazionali a bassa qualificazione nei territori caratterizzati da una forte presenza di lavoratori stranieri. Analogamente, una maggiore incidenza di dipendenti in imprese *hightech* e *knowledgeintensive* è positivamente associata alla probabilità di non rientro: ciò può indicare un disallineamento tra la struttura della domanda locale – orientata verso competenze elevate – e il profilo medio degli individui che rilasciano la DID, potenzialmente meno qualificati.

Più articolata è l'interpretazione degli indicatori di autoimpiego. Una maggiore quota complessiva di *indipendenti per 100 residenti* presenta un coefficiente negativo, suggerendo che nei territori in cui l'autoimpiego costituisce una componente strutturale consolidata del tessuto produttivo il rischio di esclusione dal mercato del lavoro risulti inferiore. Al contrario, un'elevata incidenza di *indipendenti tra i giovani (25-34 anni)* è positivamente associata alla probabilità di non reinserimento, indicando che quando l'autoimpiego è concentrato nelle coorti più giovani esso può riflettere difficoltà di accesso al lavoro dipendente piuttosto che una scelta imprenditoriale consolidata. In questo senso, l'autoimpiego appare 'protettivo' quando esprime una struttura economica matura, ma può essere associato a maggiore vulnerabilità quando rappresenta una strategia residuale di ingresso nel mercato del lavoro.

Si precisa che l'interpretazione dei coefficienti delle variabili di contesto è intrinsecamente complessa e non può essere intesa in senso causale, trattandosi di associazioni stimate in un modello con finalità predittive su dati osservazionali. Nel complesso, i risultati confermano che il reinserimento occupazionale non dipende esclusivamente dalle caratteristiche individuali, ma è fortemente influenzato dalla configurazione produttiva e demografica dei mercati locali. Ne discende una chiara implicazione di policy: gli strumenti di profiling dovrebbero incorporare in modo sistematico indicatori territoriali, così da orientare l'allocazione delle politiche attive non solo in funzione del profilo individuale, ma anche delle opportunità e dei vincoli propri dei diversi contesti locali. Un approccio multilivello – nazionale, regionale e subregionale (SLL) – appare pertanto essenziale per garantire interventi più efficaci e coerenti con le specificità dei mercati del lavoro.

4. Approcci predittivi avanzati: Random Forest e Boosting

Poiché l'obiettivo dell'analisi non è soltanto interpretativo ma anche operativo – ossia volto alla costruzione di uno strumento di profilazione utile a orientare le politiche attive – l'attenzione si estende oltre la regressione logistica verso approcci maggiormente orientati alla previsione. In questa prospettiva, accanto al modello logistico sono stati stimati modelli di machine learning, in particolare Random Forest e Boosting, con l'obiettivo di massimizzare la capacità predittiva anche a costo di una minore interpretabilità.

Un primo elemento che emerge dal confronto riguarda il livello complessivo di accuratezza (*accuracy*), intesa come quota di osservazioni correttamente classificate dal modello rispetto al totale dei casi considerati (tabella 3). Il modello Anpal stimato su dati nazionali presenta un'accuratezza pari al 67,3%, mentre i modelli stimati su dati piemontesi (logistico, con variabili di contesto, Random Forest e Boosting) si collocano su valori compresi tra il 63% e il 64,5%. Tale differenza può essere ricondotta a diversi fattori. In primo luogo, la diversa definizione della variabile dipendente: il modello Anpal considera un orizzonte temporale di 12 mesi, mentre nel caso piemontese l'esito è osservato a 6 mesi dal rilascio della DID. Un orizzonte più lungo potrebbe intercettare condizioni di disoccupazione più strutturate e persistenti, potenzialmente più facili da prevedere. Al contrario, dinamiche di breve periodo sono maggiormente influenzate da fattori congiunturali e dunque meno facilmente anticipabili.

Un ulteriore elemento riguarda la diversa disponibilità informativa. Alcune variabili incluse nel modello nazionale – in particolare quelle relative allo svolgimento di tirocini e alla qualifica professionale prevalente nei 24 mesi precedenti la DID – non trovano corrispondenza nei dati regionali e non sono quindi state incluse nel modello piemontese. Si tratta di informazioni potenzialmente rilevanti per cogliere aspetti della storia lavorativa recente, la cui assenza può incidere sulla capacità predittiva complessiva. Un'ulteriore differenza riguarda il trattamento dell'età. Nel modello nazionale essa è specificata come variabile numerica ed è oggetto di trasformazioni funzionali, consentendo di stimare in modo flessibile eventuali effetti non lineari. Nei dati piemontesi, invece, l'età è disponibile solo in forma categoriale, con conseguente perdita di informazione. La discretizzazione implica infatti che individui appartenenti alla stessa classe siano trattati in modo omogeneo e non consente di intercettare variazioni marginali interne all'intervallo considerato, incidendo potenzialmente sulla performance predittiva.

Se si considera la *sensibilità* (*sensitivity*) – ovvero la quota di soggetti rimasti disoccupati a sei mesi dal rilascio della DID e correttamente identificati

Tabella 3. Performance predittiva dei modelli di profiling

Metodo	Accuracy out-of-sample	Sensitivity	Specificity
Modello dati nazionali	67,63%	55,83%	76,76%
Modello dati piemontesi (senza var. di contesto)	63,56%	41,63%	80,32%
Modello dati piemontesi con denominazione SLL	64,12%	43,76%	79,68%
Modello dati piemontesi Best Subset Selection	64,14%	43,74%	79,72%
Modello dati piemontesi Forward Stepwise Selection	64,10%	43,71%	79,68%
Modello dati piemontesi Random Forest	64,00%	45,76%	77,93%
Modello dati piemontesi Boosting	64,50%	47,57%	77,44%

Nota: Per la definizione della variabile dipendente e del campione si rimanda alla nota della tabella 1. L'accuratezza misura la quota di osservazioni correttamente classificate; la sensibilità la quota di soggetti non occupati correttamente identificati come Weaker; la specificità la quota di soggetti Work Ready correttamente identificati.
Fonte: elaborazioni degli Autori per il modello piemontese; Anpal (2022) per il modello nazionale

dal modello – il modello Anpal raggiunge un valore del 55,8%. Ciò significa che poco più della metà dei soggetti effettivamente in stato di disoccupazione di lungo periodo viene classificata correttamente come "Weaker forte" (Anpal 2022). Pur rappresentando un risultato rilevante, questa performance evidenzia anche un ampio margine di miglioramento, in particolare se l'obiettivo è intercettare tempestivamente i soggetti che richiedono misure di sostegno più strutturate e personalizzate.

Nei modelli stimati su dati piemontesi la sensibilità è più bassa, in linea con la minore accuratezza complessiva. Il modello logistico più simile a quello Anpal a livello di predittori (senza variabili di contesto) presenta una sensibilità del 41,6%. Tuttavia, già l'introduzione della sola variabile di appartenenza all'SLL determina un incremento di circa due punti percentuali (sensibilità pari al 43,8%), pari a un incremento di circa il 5%. Un miglioramento analogo si osserva includendo le variabili di contesto selezionate a livello di SLL. Questo risultato suggerisce che la dimensione territoriale, anche in forma sintetica, contribuisce in modo significativo alla capacità di intercettare i soggetti più fragili.

Il passaggio a modelli di machine learning rafforza ulteriormente tale evidenza. Il modello Random Forest ottiene una sensibilità del 45,8%, mentre il Boosting raggiunge il 47,6%, con un incremento di circa il 14% rispetto al modello logistico senza variabili di contesto. Per questi ultimi modelli l'aumento dell'accuratezza complessiva è invece più contenuto. Considerando che, in un'ottica di policy, può essere prioritario ridurre il numero di soggetti fragili non riconosciuti, il miglioramento della sensibilità rappresenta un risultato particolarmente rilevante. In questa prospettiva, l'adozione di strumenti predittivi più flessibili rispetto al modello logistico tradizionale potrebbe rafforzare l'efficacia del sistema di profilazione.

Per quanto riguarda la *specificità* (*specificity*) – ovvero la quota di soggetti "Work-Ready" correttamente identificati dal modello – i modelli piemontesi mostrano valori generalmente superiori rispetto al modello nazionale Anpal. Il modello logistico più semplice, senza variabili di contesto, supera l'80% di specificità. Questo aspetto non è secondario: una buona capacità di riconoscere i soggetti "Work-Ready" consente di evitare l'allocazione inefficiente di risorse per accompagnare individui che presentano già elevate probabilità di reinserimento autonomo.

Nel complesso, i risultati mostrano un trade-off tra interpretabilità e performance predittiva. Tale aspetto assume particolare rilevanza nel contesto delle politiche attive del lavoro, dove gli strumenti di profiling orientano decisioni che riguardano l'impiego di risorse pubbliche e l'individuazione dei percorsi di accompagnamento al lavoro. In questo contesto, la trasparenza del modello rappresenta una caratteristica importante, poiché consente di ricostruire e comunicare i fattori associati al rischio di mancato reinserimento occupazionale. Il modello logistico rimane quindi uno strumento trasparente e facilmente comunicabile, la cui performance

predittiva migliora ulteriormente quando integra variabili di contesto territoriale. L'adozione di algoritmi di machine learning, come Random Forest e Boosting, consente invece di migliorare la capacità di intercettare i soggetti a maggiore rischio di esclusione prolungata dal mercato del lavoro. In un sistema di politiche attive orientato all'efficacia e alla personalizzazione degli interventi, tale miglioramento può tradursi in un impatto concreto sull'allocazione delle risorse e sulla qualità dei servizi offerti.

Conclusioni

Il presente lavoro ha esplorato le potenzialità dell'integrazione tra dati amministrativi individuali e informazioni territoriali per il profiling delle persone disponibili al lavoro, con l'obiettivo di migliorare la capacità predittiva dei modelli attualmente utilizzati a livello nazionale. L'analisi, condotta sui dati piemontesi della Scheda anagrafico-professionale (SAP), ha preso avvio dalla replica del modello logistico Anpal, confermando le principali regolarità empiriche già note, ed evidenziando al contempo differenze significative nella rilevanza dei vari predittori dello stato di disoccupazione, riconducibili alle specificità strutturali del mercato del lavoro regionale.

L'estensione del modello con variabili di contesto riferite ai Sistemi locali del lavoro (SLL) evidenzia il ruolo rilevante delle condizioni socio-economiche locali nella probabilità di mancato reinserimento, al netto delle caratteristiche individuali. Indicatori quali la quota di personale con cittadinanza non italiana, l'incidenza di dipendenti in imprese ad alta tecnologia e la struttura dell'autoimpiego contribuiscono a spiegare variazioni territoriali nella vulnerabilità delle persone disoccupate. L'utilizzo di tecniche di selezione delle variabili ha consentito di individuare un insieme ristretto di indicatori territoriali associati a un miglioramento della capacità predittiva del modello.

Il ricorso a modelli di machine learning – Random Forest e Boosting – ha ulteriormente incrementato la capacità di intercettare i soggetti a maggior rischio di esclusione prolungata dal mercato del lavoro (sensibilità). In un'ottica operativa, questo miglioramento può tradursi in una più efficace allocazione delle risorse pubbliche, orientando le politiche attive verso i beneficiari che necessitano di interventi più strutturati e personalizzati, da bilanciare però con la minore interpretabilità e trasparenza di questi modelli rispetto a modelli statistici tradizionali.

Nel complesso, i risultati suggeriscono che i sistemi di profiling non possono prescindere dalla dimensione territoriale. Un approccio multilivello – nazionale, regionale e subregionale – appare essenziale per garantire che gli strumenti di profilazione siano sensibili alle specificità dei mercati locali, contribuendo così a una presa in carico più mirata ed efficiente. Le evidenze emerse supportano la tesi secondo cui l'integrazione sistematica di indicatori di contesto nei modelli di profilazione può rafforzare il supporto alle decisioni dei Servizi per l'impiego, in coerenza con l'obiettivo di personalizzazione degli interventi proposto dal Programma GOL.

I risultati non evidenziano una maggiore qualità informativa dei dati piemontesi rispetto a quelli nazionali, che risultano anzi più completi, come mostrato in Appendice. Il miglioramento delle performance predittive è riconducibile all'introduzione di variabili territoriali (SLL), più che a differenze nella struttura dei dati individuali. Ne consegue che un'estensione del modello nazionale che includa analoghi indicatori di contesto appare tecnicamente realizzabile.

Al contempo, il confronto tra i coefficienti dei modelli nazionale e regionale suggerisce che l'effetto delle caratteristiche individuali possa variare a seconda del contesto territoriale, aprendo la possibilità di sviluppare specificazioni regionali del modello. Questa ipotesi merita ulteriori approfondimenti, anche alla luce dell'elevata numerosità campionaria disponibile a livello nazionale.

Il lavoro presenta alcune limitazioni. In primo luogo, la diversa finestra temporale utilizzata per definire l'esito (6 mesi su dati piemontesi, 12 mesi nel modello Anpal su dati nazionali) limita la comparabilità diretta delle performance. In secondo luogo, l'impossibilità di disporre di alcune variabili presenti nel modello nazionale (tirocini, qualifica professionale pregressa) e la differente specificazione dell'età potrebbero aver influenzato i risultati.

Ciononostante, lo studio fornisce evidenze empiriche a sostegno di un'evoluzione del profiling verso modelli ibridi, capaci di integrare informazioni individuali e territoriali con tecniche statistiche flessibili, senza rinunciare a un'adeguata trasparenza. Future ricerche potrebbero estendere l'analisi ad altre regioni italiane, adottando un disegno comparativo che consenta di generalizzare i risultati.

Appendice

La tabella A1 riporta i coefficienti stimati per i due modelli (nazionale Anpal (2022) e regionale senza variabili di contesto – dati forniti da APL - Agenzia Piemonte Lavoro), relativamente alle variabili individuali comuni. I modelli stimano la probabilità di mancato reinserimento lavorativo successivo alla sottoscrizione della DID. La variabile dipendente è pari a 1 se l'individuo non risulta occupato dopo la DID, 0

altrimenti. Nel modello piemontese l'esito è osservato a 6 mesi dalla DID (N = 163433 DID sottoscritte in Piemonte nel 2022); nel modello nazionale Anpal l'esito è osservato a 12 mesi dalla DID (circa 3,4 milioni di DID sottoscritte in Italia nel 2018-2019).

Nella tabella emergono anche le differenze tra le variabili esplicative considerate, tenendo conto che, per il modello piemontese, si è cercato di ricondurre – per quanto possibile – le variabili regionali a quelle del modello Anpal. Si ricorda inoltre che nel modello piemontese mancano completamente le informazioni relative alla partecipazione a tirocini e alla qualifica professionale prevalente nei 24 mesi precedenti la DID, variabili invece presenti nel modello nazionale.

Le categorie di riferimento sono indicate tra parentesi. Gli asterischi denotano i livelli di significatività statistica (*p<0,10; **p<0,05; ***p<0,01).

Tabella A1. Coefficienti completi del modello nazionale (Anpal) e del modello piemontese (APL)

ANPAL Variabile (cat. rif.)	APL Variabile (cat. rif.)	ANPAL modalità	APL modalità	ANPAL Coef.	APL Coef.
Genere		Donna	Donna	0,357***	0,023
Classi Età (cat. rif. inferiore a 24 anni)		Da 25 a 29 anni	Da 25 a 29 anni	0,061***	0,060**
		Da 30 a 39 anni	Da 30 a 39 anni	0,310***	0,160***
		Da 40 a 49 anni	Da 40 a 49 anni	0,483***	0,390***
		Over 50	Over 50	0,451***	0,847***
Interazione donna e classe d'età (cat. rif. uomo con età inferiore a 24 anni)		Donna - 25 a 29 anni	Donna - 25 a 29 anni	0,083***	0,042
		Donna - 30 a 39 anni	Donna - 30 a 39 anni	0,123***	0,045
		Donna - 40 a 49 anni	Donna - 40 a 49 anni	−0,100***	−0,162***
		Donna - Over 50 anni	Donna - Over 50 anni	−0,130***	−0,209***
Durata presenza in Italia (cat. rif. Cittadinanza italiana)		Fino a 12 mesi	Fino a 12 mesi	−0,111***	−0,113***
		Da 1 a 2 anni	Da 1 a 2 anni	−0,074***	−0,292***
		Oltre 2 anni	Oltre 2 anni	0,072***	−0,067***
		Nato in Italia	Nato in Italia	0,083*	1,194***
Stato occupazionale alla firma		Occupato alla DID	Occupato alla DID	0,200***	0,144***
Precedente esperienza di lavoro		Ha lavorato nei 12 mesi pre DID	Ha lavorato nei 6 mesi pre DID	−0,322***	−1,842***
Interazione classe d'età e precedente esperienza di lavoro nei 12 mesi pre DID	Interazione classe d'età e precedente esperienza di lavoro nei 6 mesi pre DID	Precedente esperienza - 25 a 29 anni	Precedente esperienza - 25 a 29 anni	0,085***	−0,001
		Precedente esperienza - 30 a 39 anni	Precedente esperienza - 30 a 39 anni	−0,125***	−0,056
		Precedente esperienza - 40 a 49 anni	Precedente esperienza - 40 a 49 anni	−0,498***	−0,119***
		Precedente esperienza - over 50 anni	Precedente esperienza - over 50 anni	−0,696***	−0,364***

segue

segue tabella A1

Titolo di studio (cat. rif. Fino licenza elementare)	Titolo di studio (cat. rif. Nessun titolo)	Licenza elementare	Licenza elementare	-	−0,226**
		Licenza media	Scuola dell'obbligo	−0,073***	−0,177*
		Liceo		−0,024***	
		Qualifica professionale	Istruzione professionale	−0,253***	−0,247**
		Istituto professionale		−0,281***	
		Istituto tecnico		−0,188***	
Diploma o Triennale		Diploma	Diploma	−0,175***	−0,101
		Scienze umanistiche	Laurea I o II livello	−0,356***	0,015
		Scienze sociali		−0,200***	
		Scienze della salute		−0,514***	
		Ingegneria, informatica e trasporti		−0,547***	
		Altro diploma o triennale		−0,239***	
Laurea magistrale, specialistica, vecchio ordinamento		Scienze umanistiche	Corsi post laurea	−0,558***	0,127
		Scienze sociali		−0,257***	
		Scienze della salute		−0,666***	
		Ingegneria, informatica e trasporti		−0,741***	
		Scienze naturali		−0,571***	
		Architettura		−0,042**	
		Altra laurea magistrale, specialistica, vecchio ordinamento		−0,348***	
Patente		Possedere una patente	Possedere una patente	−0,189***	−0,595***
Condizione occupazionale prevalente anno precedente (cat. rif. Altro inattivo)		Occupato	Occupato	−0,007	−0,410***
		In cerca di nuova occupazione	In cerca di nuova occupazione	−0,331***	−0,322***
		In cerca di prima occupazione	In cerca di prima occupazione	0,066***	−0,012
		Studente	Studente	−0,141***	−0,056*
Settore prevalente 24 mesi precedenti la firma della DID (cat. rif. Non ha lavorato)	Settore prevalente 6 mesi precedenti la firma della DID (cat. rif. Non ha lavorato)	Agricoltura	Agricoltura	−0,823***	−0,410***
		Industria in senso stretto	Industria	0,023**	−0,454***
		Costruzioni	Costruzioni	−0,216***	−0,003
		Commercio	Commercio	0,121***	−0,083
		Alloggio e ristorazione	Alloggio e ristorazione	−0,451***	−0,235***
		Trasporto, immagazzinaggio e altri servizi	Servizi	−0,140***	−0,364***

segue

segue tabella A1

Interazione donna e settore prevalente 24 mesi precedenti (cat. rif. uomo - non ha lavorato)	Interazione donna e settore prevalente 6 mesi precedenti (cat. rif. uomo - non ha lavorato)	Donna - Agricoltura	Donna - Agricoltura	−0,197***	0,157
		Donna - Industria in senso stretto	Donna - Industria	−0,088***	−0,208***
		Donna - Costruzioni	Donna - Costruzioni	0,619***	0,419***
		Donna - Commercio	Donna - Commercio	−0,134***	−0,076
		Donna - Alloggio e ristorazione	Donna - Alloggio e ristorazione	−0,051***	0,006
		Donna - Trasporto, immagazzinaggio e altri servizi	Donna - Servizi	0,014*	0,022
Numero datori cambiati nei 24 mesi precedenti la firma della DID (cat. rif. Non ha lavorato nel periodo)	Numero contratti avuti nei 6 mesi precedenti la firma della DID (cat. rif. Nessuno)	1 datore di lavoro	1 lavoro	0,863***	1,458***
		2 datori di lavoro	2-5 lavori	0,497***	0,792***
		Più di 2 datori di lavoro	5-10 lavori	0,000	0,000
			10 o più lavori		−0,402*
Impegno familiare		Famiglia con figli	Presenza di figli a carico	−0,058***	−0,055**
Interazione donna e impegno familiare		Donna con figli	Donna - Presenza di figli a carico	0,373***	0,096***
Numero componenti famiglia (cat. rif. Unico componente)		2	2	0,020***	0,003
		3	3	−0,063***	−0,043**
		4	4	−0,138***	−0,032*
		5	5	−0,102***	−0,002
		Più di 5	Più di 5	0,009	0,027
Costante				1,650***	1,076***

Fonte: stime degli Autori per il modello piemontese; coefficienti riportati in Anpal (anno) per il modello nazionale

Bibliografia

Agenzia per l'Italia digitale (AgID) (2021), *Conservazione dei documenti digitali*, Roma, Agid

Akaike H. (1974), A new look at the statistical model identification, *IEEE Transactions on Automatic Control*, 19, n.6, pp.716-723

Bazzoli M., Marzadro S., Trivellato U. (2018), *Un'esperienza pilota di integrazione di dati amministrativi e di survey per l'analisi dell'evoluzione delle storie lavorative dei giovani*, FBK-IRVAPP Working Papers n.2018-01, Trento, Fondazione Bruno Kessler-IRVAPP

Bollens J. (2022), *Evaluatie van de effectiviteit van diverse activerende arbeidsmarkt-maatregelen bij VDAB*, Brussel, VDAB

Burnham K.P., Anderson D. R. (2002), *Model Selection and Multimodel Inference: A Practical Information-Theoretic Approach*, New York, Springer, 2nd ed

Chetty R. (2012), *Time trends in the use of administrative data for empirical research,* 34th Annual NBER Summer Institute, Cambridge, Massachusetts

Desiere S., Langenbucher K., Struyven L. (2019), *Statistical profiling in public employment services: An international comparison*, OECD Social, Employment and Migration Working Papers n.224, OECD Publishing, Paris

Dipartimento della Funzione Pubblica (2024), *6° Piano d'Azione nazionale per il governo aperto 2024-2026*, Roma

Istat (2025), *I Sistemi locali del lavoro 2021*, Focus, Roma, Istat

Istat (2014), *Progetto ARCHIMEDE. I risultati sperimentali e le linee di sviluppo*, Istat Working Papers, Roma, Istat

Leombruni R., Ardito C. (2026), The long and winding road. The access to administrative data for research purposes in Italy, in Berton F., Paruolo P. (eds.), *Data-Driven Learning in the EU. From Administrative Sources to Public Value*, Cham, Springer

Leombruni R. (2024), Interviewing administrative records: A conceptual map for the use of big data for economic research, *Statistica Applicata. Italian Journal of Applied Statistics*, 36, n.3, pp.295-326

Loxha A., Morgandi M. (2014), *Profiling the unemployed. A review of OECD experiences and implications for emerging economies*, Social Protection and labor discussion paper n.SP1424, Washington, World Bank

O'Connell P. J., McGuinness S., Kelly E., Walsh J. (2009), *National Profiling of the Unemployed in Ireland. Research Report*, Dublin, Economic and Social Research Institute

Trivellato U. (2019), Microdata for social sciences and policy evaluation as a public good, in Crato N., Paruolo P. (eds.), *Data-Driven Policy Impact Evaluation*, Cham, Springer, pp.27-45

Monica Bosia
mbosia@agenziapiemontelavoro.it
Analista presso Agenzia Piemonte Lavoro – Servizio Monitoraggio, Studi e Ricerche. Laureata in Scienze politiche, sociali e dell'amministrazione presso l'Università del Piemonte Orientale, dal 2016 si occupa di analisi di dati amministrativi, data mining e analisi delle politiche attive del lavoro a fini di monitoraggio e rendicontazione.

Marianna Carlone
marianna.carlone@edu.unito.it
Laureata magistrale in Metodi statistici ed economici per le decisioni presso il Dipartimento di Economia e statistica "Cognetti de Martiis", Università degli Studi di Torino.

Elisa Floredan
efloredan@agenziapiemontelavoro.it
Specialista delle politiche del lavoro presso Agenzia Piemonte Lavoro – Servizio Monitoraggio, Studi e Ricerche. Laureata magistrale in Sociologia presso l'Università degli Studi di Torino, dal 2020 svolge attività di ricerca quali-quantitativa sui mercati del lavoro e sulle politiche attive territoriali, occupandosi della progettazione e della reportistica periodica per il monitoraggio e la definizione degli scenari locali del lavoro.

Francesca Prever
francesca.prever@unito.it
Borsista di ricerca presso il Dipartimento di Economia e statistica "Cognetti de Martiis", Università degli Studi di Torino.

TREMOD-Imprese: un modello di microsimulazione

Martina Bazzoli
FBK-IRVAPP

Sonia Marzadro
FBK-IRVAPP

Annalisa Tassi
FBK-IRVAPP

Carlo Fiorio
FBK-IRVAPP e Università degli Studi di Milano

L'articolo presenta TREMOD-Imprese, un modello di microsimulazione basato su dati fiscali amministrativi della Provincia autonoma di Trento per la valutazione ex ante della fiscalità d'impresa. Integrando dichiarazioni IRES, IRAP e IVA con l'anagrafe delle imprese, il modello ricostruisce basi imponibili e imposta netta e consente simulazioni statiche degli effetti. Il contributo si inserisce nel filone, ancora limitato, della microsimulazione applicata alla fiscalità d'impresa e mostra come l'integrazione di dati amministrativi possa rafforzare il supporto analitico alle decisioni di policy a livello territoriale. Sebbene le analisi presentate facciano riferimento ai dati fiscali del 2019, il modello è attualmente in fase di estensione temporale al periodo 2015–2023.

TREMOD-Imprese: A microsimulation model

The article presents TREMOD-firms, a microsimulation model based on administrative tax data from the Autonomous Province of Trento for the ex ante evaluation of corporate taxation. By integrating IRES, IRAP and VAT returns with the business register, the model reconstructs tax bases and net liabilities and enables static simulations of revenue effects. The contribution fits within the still limited literature on corporate tax microsimulation and shows how the integration of administrative data can strengthen analytical support for policy decisions at the territorial level. Even though the analyses are based on data for 2019, an update of the model to the period 2015-2023 is ongoing.

DOI: 10.53223/Sinappsi_2026-02-11

Citazione
Bazzoli M., Marzadro S., Tassi A., Fiorio C. (2026), TREMOD-Imprese: un modello di microsimulazione, *Sinappsi*, XVI, n.2, pp.161-173

Parole chiave
Dati amministrativi
Imprese
Valutazione delle politiche

Keywords
Administrative data
Enterprises
Policy evaluation

Introduzione

I dati fiscali costituiscono una risorsa di elevato valore strategico per l'analisi empirica e la progettazione delle politiche pubbliche, in particolare nell'ambito della valutazione ex ante degli interventi di fiscalità d'impresa. La loro rilevanza risiede nella copertura dell'intera platea dei contribuenti, che assicura un patrimonio informativo esteso, sistematico e aggiornato su base annuale. Tale caratteristica consente di superare i limiti delle tradizionali indagini campionarie e di realizzare analisi ad elevato livello di granularità anche su segmenti ristretti del tessuto produttivo. Al contempo, la natura censuaria della fonte permette di collocare tali segmenti all'interno del quadro complessivo del sistema delle imprese. In questo senso, i dati fiscali si prestano in modo particolarmente efficace a supportare esercizi di valutazione ex ante, permettendo di analizzare in maniera coerente gli effetti potenziali delle misure fiscali sia in termini aggregati, attraverso la stima delle variazioni di gettito, sia dal punto di vista distributivo, considerando l'eterogeneità delle imprese per dimensione, settore e localizzazione.

Questo approccio metodologico è al centro del presente contributo e costituisce la base su cui si sviluppa il progetto TREMOD-Imprese, messo a punto nel contesto della Provincia autonoma di Trento (PaT). Il progetto integra le dichiarazioni fiscali IRES, IRAP e IVA con le informazioni dell'anagrafe delle imprese e con i dati di bilancio, con l'obiettivo di costruire uno strumento conoscitivo per simulazioni sulla fiscalità

d'impresa. Il progetto viene qui presentato nei suoi presupposti analitici, nelle scelte di integrazione dei dati e nelle prime applicazioni empiriche.

La Provincia autonoma di Trento rappresenta un contesto istituzionale particolarmente adatto allo sviluppo di questo tipo di analisi. I margini di autonomia in ambito fiscale e regolatorio consentono infatti di osservare e simulare interventi di policy che incidono direttamente su basi imponibili e aliquote, offrendo un banco di prova rilevante per l'applicazione di strumenti di analisi fondati su dati amministrativi integrati. In questo contributo, il contesto trentino non è utilizzato come caso studio in senso descrittivo, ma come laboratorio istituzionale per discutere in modo esplicito le scelte metodologiche, i vincoli informativi e le potenzialità analitiche associate all'utilizzo dei dati fiscali amministrativi nella valutazione ex ante.

Dal punto di vista metodologico, il contributo si colloca all'intersezione tra la letteratura sulla microsimulazione fiscale e quella sul federalismo fiscale sub-nazionale. Rispetto ai modelli nazionali esistenti, TREMOD-Imprese si distingue per tre elementi. In primo luogo, opera a livello sub-nazionale in un contesto caratterizzato da autonomia finanziaria, consentendo di analizzare il nesso tra leva fiscale e bilancio territoriale. In secondo luogo, integra sistematicamente dichiarazioni fiscali e anagrafe delle imprese, permettendo un'analisi congiunta di basi imponibili, caratteristiche strutturali delle imprese e distribuzione territoriale del gettito. In terzo luogo, si configura come uno strumento orientato esplicitamente al *policy design* locale, rendendo operativa la possibilità di valutazioni ex ante in un contesto in cui una parte rilevante del gettito rimane sul territorio.

Il contributo è strutturato come segue. La sezione successiva inquadra il ruolo dei dati fiscali amministrativi nella valutazione ex ante delle politiche pubbliche alla luce della letteratura esistente. Segue una descrizione delle fonti amministrative utilizzate in TREMOD-Imprese, con particolare attenzione alle modalità di integrazione dei dati e ai vincoli informativi rilevanti. Vengono quindi presentate le principali evidenze descrittive sulla popolazione di imprese osservata e le prime applicazioni in termini di simulazioni fiscali. L'ultima sezione discute i limiti dell'analisi e le prospettive di sviluppo del progetto.

1. Dati fiscali amministrativi come strumento di analisi ex ante

La letteratura che impiega dati fiscali amministrativi per l'analisi ex ante delle politiche pubbliche evidenzia due elementi principali. In primo luogo, gli esercizi di valutazione ex ante rappresentano un filone relativamente più ristretto rispetto all'ampia produzione di studi basati su valutazioni ex-post e su stime causali (Crato e Paruolo 2019). Quando presenti, tali esercizi sono generalmente motivati da esigenze di *policy design* – come il confronto tra alternative di intervento o la simulazione di riforme – e si concentrano prevalentemente sugli effetti contabili e distributivi delle misure fiscali, quasi sempre in assenza di risposte comportamentali.

Un secondo elemento chiave riguarda l'ambito di applicazione. La maggior parte delle analisi ex ante basate su dati fiscali amministrativi si concentra infatti sui sistemi di welfare e sulla tassazione individuale, mentre le applicazioni riferite alla fiscalità d'impresa risultano decisamente più limitate. Nel primo ambito, Euromod (Sutherland e Figari 2013; Figari *et al.* 2019) rappresenta uno dei principali riferimenti metodologici e applicativi a livello europeo. Il modello consente di simulare ex ante riforme fiscali e trasferimenti monetari in modo armonizzato tra Paesi, basandosi prevalentemente su dati di indagine campionaria integrati con informazioni amministrative e con una rappresentazione dettagliata delle regole fiscali. Euromod ha dato origine a una vasta letteratura ed è ampiamente utilizzato come strumento di supporto al *policy design*, ma il suo campo di applicazione rimane in larga misura circoscritto agli individui e alle famiglie. Esperienze analoghe sono rintracciabili anche a livello sub-nazionale (per l'Italia, Azzolini *et al*. 2017; Figari *et al.* 2019; Maitino *et al.* 2017), anch'esse prevalentemente orientate alla tassazione personale e alle politiche redistributive.

Le applicazioni ex ante alla fiscalità d'impresa sono invece rare e concentrate in pochi modelli nazionali. In Italia, le principali esperienze sono il modello Istat-MATIS, basato sulle dichiarazioni IRES e IRAP e orientato alla stima del carico fiscale effettivo in assenza di risposte comportamentali (Caiumi e Di Biagio 2016)[1]; il modello CITSIM-DF del Ministero dell'Economia e delle finanze, che utilizza dati IRES e di bilancio e mira allo sviluppo di un modulo comportamentale (Bellucci *et al.* 2023); e il modello MEDITA dell'Ufficio Parlamentare di bilancio che integra dati da dichiarazioni IRAP e IRES a dati di bilancio con potenziali future estensioni a informazioni sui contributi previdenziali (Gastaldi 2019). In ambito internazionale, il progetto tedesco ZEW TaxCoMM si basa prevalentemente su dati di bilancio, con conseguenti limitazioni nella

1 Si veda ad esempio Istat (2024), Comunicato Stampa, *Gli effetti distributivi dei provvedimenti fiscali sulle imprese,* 5 luglio, per un'applicazione di Istat-MATIS alla simulazione dei nuovi provvedimenti fiscali in vigore nel 2024 (*nowcasting*).

simulazione delle basi imponibili (Finke *et al.* 2010), mentre Shahnazarian (2011), elabora un modello dinamico per le società per azioni in Svezia[2]. A livello provinciale o regionale, non ci risultano esperienze consolidate di modelli di utilizzo dei dati amministrativi per la valutazione ex ante della fiscalità d'impresa.

È proprio in questo spazio ancora poco esplorato che si colloca il presente contributo, che estende l'utilizzo dei dati fiscali amministrativi e degli approcci di simulazione ex ante al lato delle imprese in un contesto sub-nazionale caratterizzato da autonomia fiscale.

Un ultimo aspetto rilevante relativamente all'utilizzo di dati amministrativi per l'analisi delle politiche pubbliche riguarda le condizioni di accesso ai dati che richiedono, infatti, assetti istituzionali e organizzativi tali da poter garantire la sicurezza delle informazioni e la tutela della riservatezza. Nel caso di TREMOD-Imprese, l'interazione strutturata tra il Dipartimento Affari Finanziari della Provincia autonoma di Trento, l'Istituto di Statistica provinciale (Ispat) e l'Istituto per la Ricerca Valutativa sulle Politiche Pubbliche della Fondazione Bruno Kessler (FBK-IRVAPP) ha reso concretamente utilizzabili basi informative complesse per finalità di *policy design*, contribuendo non solo alla produzione di evidenze empiriche, ma anche alla definizione di condizioni istituzionali potenzialmente rilevanti per esperienze analoghe in altri contesti.

2. Le imposte d'impresa

Le dichiarazioni fiscali raccolgono un insieme ampio e strutturato di informazioni sulle caratteristiche economiche delle imprese, sulla loro localizzazione territoriale, sulla dimensione dell'attività e sulle principali grandezze rilevanti ai fini del calcolo dell'imposta. In questa prospettiva, il sistema tributario può essere letto anche come un'infrastruttura informativa utile per supportare l'analisi e il disegno delle politiche pubbliche.

Nel contesto italiano, IRES, IRAP e IVA costituiscono le principali imposte di riferimento per l'analisi della fiscalità d'impresa. Ciascuna di esse riflette una diversa dimensione dell'attività economica: il reddito prodotto, il valore della produzione e il valore aggiunto generato nelle transazioni. La compresenza di queste tre imposte consente quindi di osservare le imprese da prospettive complementari e di ricostruire, attraverso l'integrazione delle relative dichiarazioni, un quadro informativo particolarmente ricco.

Queste imposte presentano caratteristiche diverse ma fortemente complementari. Le dichiarazioni IRES offrono informazioni dettagliate sulla formazione del reddito imponibile delle società di capitali e degli enti non commerciali; le dichiarazioni IRAP consentono di cogliere la dimensione territoriale (provinciale o regionale) dell'attività produttiva, risultando particolarmente rilevante in contesti caratterizzati da autonomia fiscale e più in generale per le finanze pubbliche territoriali; le dichiarazioni IVA forniscono infine una misura del volume delle transazioni e del valore aggiunto generato, con una copertura estesa anche alle imprese di minori dimensioni.

La scelta di focalizzarsi su queste tre imposte risponde dunque a una duplice esigenza. Da un lato, esse rappresentano una quota rilevante del prelievo complessivo sulle imprese e costituiscono strumenti centrali per la politica fiscale. Dall'altro lato, le relative dichiarazioni amministrative offrono un patrimonio informativo che, se integrato e analizzato congiuntamente, consente di sviluppare analisi descrittive, simulazioni ex ante, e studi d'impatto ex post utili a informare il decisore pubblico, in particolare a livello sub-nazionale.

Nel contesto della Provincia autonoma di Trento, tale potenziale informativo assume una rilevanza particolare in ragione dell'autonomia finanziaria riconosciuta dallo statuto speciale. La possibilità di intervenire su alcune leve fiscali e, soprattutto, la permanenza sul territorio di una parte del gettito rendono le simulazioni non soltanto esercizi conoscitivi, ma strumenti operativi potenzialmente di supporto alla programmazione di bilancio e al disegno delle politiche economiche.

Prima di procedere alla descrizione analitica delle singole imposte, è pertanto opportuno chiarire la relazione tra il gettito osservabile attraverso le dichiarazioni fiscali delle imprese localizzate sul territorio e le effettive entrate del bilancio provinciale trentino. A differenza dell'IRAP, il cui gettito è di competenza diretta della provincia, il gettito IRES e IVA è determinato secondo criteri di riparto stabiliti a livello nazionale e non coincide con il gettito effettivamente raccolto presso i contribuenti trentini. In particolare, la quota di IRES attribuita alla Provincia autonoma di Trento è calcolata secondo un criterio suppletivo basato sul rapporto tra PIL locale e PIL nazionale, applicato al 90% dell'incasso IRES complessivo a livello nazionale. Analogamente, la quota di IVA spettante è determinata in base al rapporto tra consumi familiari locali e consumi

2 Sono inoltre in fase di sviluppo dei modelli di microsimulazione d'impresa anche a livello UE: https://joint-research-centre.ec.europa.eu/projects-and-activities/fiscal-policy-analysis/corporate-taxation/corporate-tax-microsimulation-model-direct_en.

familiari nazionali[3], applicato al 90% dell'incasso netto IVA nazionale.

Alla luce di queste considerazioni, di seguito si presenta una sintesi delle principali caratteristiche di IRES, IRAP e IVA, finalizzata a chiarire quali elementi del sistema fiscale siano osservabili attraverso le dichiarazioni amministrative e in che modo tali informazioni entrino nella costruzione della base dati e nelle analisi successive. La descrizione non ha finalità normative, ma è funzionale a comprendere la struttura delle informazioni disponibili, le differenze tra le tre imposte in termini di soggetti dichiaranti, base imponibile e dimensione territoriale, e le implicazioni di tali differenze per l'analisi empirica e per le simulazioni di policy.

IRES

L'IRES è un'imposta sui redditi generati dalle società. Sono principalmente soggetti a questa imposta le società di capitali, gli enti non commerciali e gli enti pubblici e privati residenti in Italia, assieme alle società e gli enti non residenti in Italia per i redditi ivi prodotti. Le società di persone, e in particolare le persone fisiche esercenti un'attività d'impresa, arti, o professioni non sono, invece, soggette a questa imposta, mentre sono soggette all'IRPEF[4]. La base imponibile è quindi rappresentata dai redditi (di qualsiasi fonte e tipologia) prodotti dalle società durante il periodo d'imposta. Per le società di capitali, la base imponibile viene definita a partire dall'utile o dalla perdita riportati in bilancio, e aggiustata in base alla normativa fiscale, tenendo conto delle cosiddette variazioni in aumento o diminuzione degli elementi attivi o passivi del reddito. L'aliquota ordinaria è del 24%, ma è possibile anche l'applicazione di un'aliquota ridotta del 12% ad esempio per società sportive dilettantistiche ed enti ospedalieri[5]. Per gli enti non commerciali, i redditi sono costituiti dalla somma dei redditi provenienti da diverse fonti (ad esempio redditi fondiari, di capitale, d'impresa), in maniera simile alle persone fisiche, e ad esclusione dei redditi esenti.

IRAP

L'IRAP è un'imposta diretta sulle attività produttive; è dovuta cioè per l'esercizio di attività che hanno come scopo lo scambio di beni o la prestazione di servizi. Sono soggetti all'IRAP sia le società di capitali e di persone, che gli enti non commerciali e gli organi e amministrazioni dello Stato[6]. Anche le società e gli enti non residenti in Italia sono soggetti all'imposta se vi hanno svolto attività per un periodo non inferiore a tre mesi mediante stabile organizzazione. I soggetti che rientrano nel regime forfettario non sono, invece, soggetti all'IRAP[7].

La base imponibile viene definita a partire dal valore della produzione derivata dall'attività svolta durante il periodo d'imposta e viene aggiustata in base alla normativa fiscale[8]. L'elemento distintivo dell'IRAP è la sua dimensione territoriale: le aliquote possono variare tra territori e settori, come nel caso dei settori assicurativo e bancario. Il gettito è di competenza delle Regioni e delle Province autonome, che possono modulare le aliquote entro i limiti fissati dalla legge. Nel caso della Provincia autonoma di Trento, l'aliquota ordinaria è pari al 2,68%, con esenzioni temporanee per le nuove attività (per i primi cinque anni di attività) e aliquote agevolate (dell'1,68%) legate all'occupazione, oltre a maggiorazioni per specifici settori (ad esempio, esercizi con apparecchi da gioco d'azzardo, banche, e assicurazioni che arrivano fino al 6,82%). Per calcolare l'imposta dovuta, i soggetti devono tener conto della distribuzione dell'attività produttiva nelle varie regioni e province autonome. La ripartizione dell'attività produttiva tra regioni e province autonome avviene in misura proporzionale all'ammontare delle retribuzioni, dei compensi e degli utili spettanti, rispettivamente, al personale dipendente e agli associati, ubicati nel territorio della regione (o provincia autonoma) e operanti per almeno tre mesi.

IVA

L'IVA invece è un'imposta indiretta, basata sul valore aggiunto e dovuta sulla cessione di beni e sulle prestazioni di servizi. È dovuta dai soggetti che esercitano attività d'impresa, arti o professioni e, in ultima istanza, grava sui consumatori finali. Le imprese svolgono un ruolo di intermediazione tra il consumatore e l'erario, addebitando l'imposta in fattura e detraendo l'IVA assolta sugli acquisti

3 L'Indagine sulle spese delle famiglie (Istat) rappresenta la fonte dei consumi familiari a livello sub-nazionale (regionale o provinciale) (Santoro 2003).

4 Regole ancora diverse si applicano a coloro che rientrano nel regime forfettario.

5 In casi eccezionali (nel nostro campione) si applica l'aliquota maggiorata del 27,5% (per soggetti che producono redditi derivanti da attività svolte sulla base di concessioni autostradali, aeroportuali, e portuali).

6 Le amministrazioni pubbliche non vengono considerate nell'analisi.

7 Più in generale, dal 2022 le persone fisiche non sono più soggette all'IRAP.

8 Nel periodo che analizziamo (2015-2019) non era ancora disponibile la deduzione integrale del costo dei dipendenti dal valore della produzione, inserita a partire dall'anno d'imposta 2021.

dall'imposta dovuta.

Sono soggetti passivi IVA le persone fisiche con ditta individuale, le società di persone e di capitali, nonché i soggetti non residenti con stabile organizzazione in Italia. I soggetti in regime forfettario non sono soggetti all'imposta. L'IVA viene versata su base periodica, mediante liquidazioni mensili o trimestrali, e conguagliata annualmente attraverso la dichiarazione IVA. La base imponibile è definita a partire dal volume d'affari, opportunamente rettificato secondo la normativa vigente. L'aliquota ordinaria è pari al 22%, affiancata da aliquote ridotte e da regimi di esenzione o non imponibilità per specifiche operazioni. Le più comuni riguardano ad esempio alimenti, bevande e prodotti agricoli (4%), oppure la fornitura di gas ed elettricità (10%). Esportazioni e cessioni intracomunitarie non sono imponibili, mentre prestazioni come quelle sanitarie o le attività educative e culturali sono esenti.

3. La costruzione di TREMOD-Imprese

TREMOD-Imprese si basa sull'integrazione delle informazioni provenienti dalle dichiarazioni fiscali relative alle tre imposte sopra descritte riferite all'anno d'imposta 2019[9]. In particolare, il database include:

1. le dichiarazioni IRES delle società di capitali non consolidate (Modello SC) e degli enti non commerciali (Modello ENC). Nel 2019 per IRES SC e ENC osserviamo 12.702 dichiaranti.
2. le dichiarazioni IRAP delle imprese che riportano attività produttive in Provincia autonoma di Trento. Nel 2019 osserviamo 33.705 dichiaranti.
3. le dichiarazioni annuali IVA dei soggetti fiscalmente residenti nella Provincia autonoma di Trento che possono svolgere attività nel territorio locale o nazionale, ma anche avere clienti e fornitori all'estero. In totale, si tratta di 44.588 dichiaranti.

I dati delle dichiarazioni sono elaborati col software statistico Stata (versione 19) e vengono utilizzati tre moduli separati per la simulazione delle tre imposte. Per ciascun dichiarante sono disponibili informazioni sulla natura giuridica, sullo stato dell'attività (in esercizio, in liquidazione o in fallimento) e sul settore economico di appartenenza, secondo la classificazione ATECO.

Per l'IRES, i dati includono informazioni sull'utile o sulla perdita, sulle variazioni fiscali in aumento o diminuzione, sulle basi imponibili, sulle aliquote applicate, sulle detrazioni e sui crediti d'imposta[10]. Per l'IRES, partendo dalla base imponibile, riusciamo a simulare, per ciascuna unità, l'imposta lorda e netta (tenendo conto delle detrazioni totali), e infine l'imposta dovuta (tenendo conto dei crediti totali).

Per l'IRAP sono disponibili informazioni sul valore della produzione, sulle deduzioni e sui crediti regionali, nonché sulle aliquote applicate, consentendo di tenere conto della dimensione territoriale dell'imposta. In questo caso, partendo dal valore della produzione possiamo simulare, per ciascuna unità, la base imponibile (tenendo conto di eventuali quote GEIE e deduzioni regionali), la relativa imposta lorda (sulla base delle aliquote date), e l'imposta netta (al netto di detrazioni e crediti regionali).

Per l'IVA, infine, sono osservabili il volume d'affari, le operazioni attive e passive e le principali componenti rilevanti per la determinazione dell'imposta dovuta. A partire dalle basi imponibili delle operazioni attive e passive soggette alle varie aliquote riusciamo a simulare l'IVA dovuta sull'attivo e l'IVA totale sul passivo. A partire poi dall'IVA a debito e detraibile, riusciamo a calcolare l'imposta dovuta o l'imposta a credito[11].

Attraverso un collegamento puntuale basato sul codice fiscale dell'impresa, le dichiarazioni fiscali sono integrate con le informazioni contenute nell'archivio anagrafico delle imprese (ASIA), che fornisce dati sulla dimensione aziendale, sul numero di addetti, sulla localizzazione della sede legale e dell'attività, nonché sulla classe di fatturato. Si noti che ASIA non copre tutte le attività economiche, escludendo, tra le altre, l'agricoltura, la silvicoltura, le istituzioni pubbliche e alcune organizzazioni non profit.

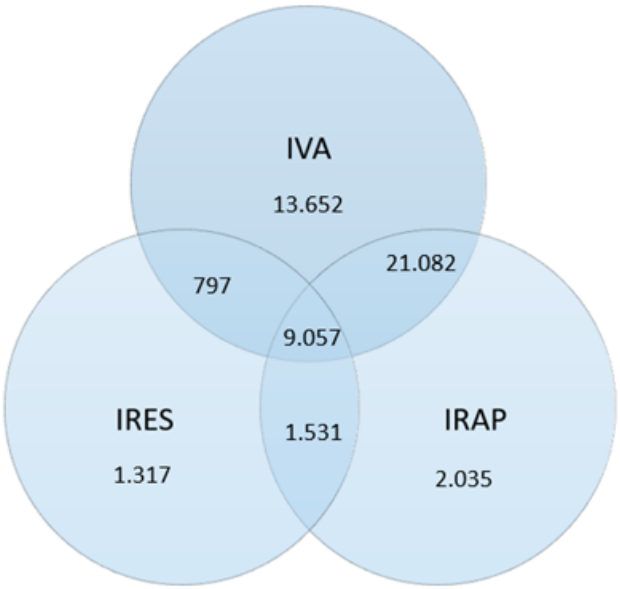
Le intersezioni tra le diverse imposte sono riportate in figura 1. Su un totale di circa 49.500 dichiaranti, una quota pari a circa il 18% (prevalentemente di società di capitali ed altri enti non commerciali) risulta soggetta simultaneamente a IRES, IRAP e IVA, mentre la maggioranza dei soggetti è assoggettata a una combinazione parziale di imposte, riflettendo l'eterogeneità delle forme giuridiche e delle attività economiche presenti

9 Per IRES e IRAP, i dati contenuti in TREMOD-imprese coprono gli anni d'imposta 2015-19 mentre per IVA solo il 2019. Per ragioni espositive il testo si riferisce al solo 2019 come anno d'imposta più recente.

10 Le informazioni acquisite sono una selezione di quelle presenti nei quadri RF, RN per le società di capitale e nei quadri RE, RF, RG e RN per gli enti non commerciali. Nella scelta sono state privilegiate quelle utili per supportare le simulazioni su possibili modifiche dell'imposta che non implicassero cambiamenti nel calcolo del reddito imponibile.

11 Non avendo dati dal quadro VJ (imposte relative a particolari tipologie di operazioni), non riusciamo a simulare l'IVA a debito (voce da dichiarazione VL1). Inoltre, non abbiamo le voci sui crediti d'imposta per poter simulare il totale IVA a credito o dovuta.

Figura 1. Numero di imprese per tipo di imposta dichiarata o versata



Fonte: elaborazione su dati dichiarazioni Agenzia delle entrate

sul territorio. Una più grossa porzione (attorno al 42%) è invece solamente soggetta a IRAP e IVA (ma non a IRES), ovvero le imposte che riguardano più soggetti, perché includono anche società di persone e persone fisiche (liberi professionisti e imprese individuali). Un'altra porzione rilevante (quasi il 28%) presenta solo dichiarazioni IVA, di cui metà sono imprese agricole (quindi esenti dall'IRAP). Le circa 2.800 imprese dichiaranti solo IRES o solo IRES e IRAP sono per oltre il 90% enti non commerciali plausibilmente esenti dall'IVA (hanno un'imponibile IRES mediano sotto i 250 euro). Le circa 2.000 imprese dichiaranti solo IRAP sono per un terzo *null filers*, e per un altro terzo di imprese con sede legale fuori dalla PaT, che però svolgono attività produttive in provincia.

4. Analisi descrittive

Le analisi descrittive presentate in questa sezione hanno l'obiettivo di fornire una prima panoramica delle caratteristiche delle imprese osservate e delle principali grandezze fiscali rilevanti ai fini di IRES, IRAP e IVA[12]. Tali evidenze consentono di descrivere la composizione della base imponibile, la distribuzione delle imprese per dimensione e l'allocazione del gettito tra diverse tipologie di soggetti, offrendo indicazioni utili per comprendere il funzionamento del sistema fiscale a livello locale.

Le statistiche riportate rappresentano un passaggio preliminare essenziale per informare il decisore pubblico sulla struttura della popolazione di riferimento e per fornire il contesto necessario all'interpretazione delle simulazioni di impatto ex ante presentate nella sezione successiva.

Le tabelle che seguono si concentrano in particolare su tre dimensioni rilevanti per la policy: (i) la distribuzione delle basi imponibili per aliquota fiscale, (ii) la dimensione delle imprese in termini di addetti, e (iii) l'allocazione del gettito tra imprese di diversa dimensione. Queste dimensioni permettono di evidenziare potenziali elementi di eterogeneità e di fornire una base informativa utile per la progettazione e la valutazione di interventi fiscali a livello provinciale.

La tabella 1 mostra come le basi imponibili relative a IRES, IRAP e IVA si distribuiscano tra le diverse aliquote applicabili, evidenziando differenze rilevanti tra le tre imposte. Nel caso dell'IRES, la base imponibile delle società di capitali risulta fortemente concentrata sull'aliquota standard, mentre per gli enti non commerciali emerge un utilizzo più frequente di aliquote inferiori. Per l'IRAP, la distribuzione riflette la coesistenza di aliquote differenziate, legate sia alla normativa nazionale sia alle scelte regionali.

Per quanto riguarda l'IVA, la base imponibile si

12 Nelle tabelle descrittive che seguono non consideriamo le imprese con base imponibile zero relativamente a IRES e IRAP e, relativamente a IVA, circa 1.100 imprese che non riportano né una base imponibile attiva né passiva, ma che, in circa metà dei casi, riportano solo IVA a credito da crediti e correzioni.

Tabella 1. Distribuzione percentuale della base imponibile per aliquota fiscale (IVA, IRES, IRAP), base imponibile complessiva in milioni di euro e numero di imprese. Anno 2019

	IRES		IRAP	IVA	
	SC	ENC		Operaz. attive	Operaz. Passive
Aliquota zero	0,0	0,0	22,3	12,5	0,0
Aliquota minore dell'aliquota standard	14,4	51,1	12,0	30,8	25,3
Aliquota standard	85,6	48,8	64,8	56,7	74,7
Aliquota maggiore dell'aliquota standard	0,0	0,0	0,8	0,0	0,0
Totale	100,0	100,0	100,0	100,0	100,0
Totale base imponibile (milioni euro)	1.036	49	4.181	27.963	24.982
N. imprese	5.321	2.571	24.919	43.481	

Nota: Si considera la base imponibile diversa da zero per IRES e IRAP, si escludono i dichiaranti che non riportano né base imponibile attiva né passiva per IVA.
Fonte: elaborazione su dati dichiarazioni Agenzia delle entrate

Tabella 2. Distribuzione percentuale delle imprese soggette a IVA, IRES e IRAP per classe di addetti. Anno 2019

Classe addetti	IRES SC[1]	IRAP	IVA
1-9 (micro)	71,2	80,5	64,1
10-49 (piccole)	19,1	8,7	5,2
50-249 (medie)	2,5	1,1	0,6
250+ (grandi)	0,4	0,2	0,1
Missing	6,8	9,5	29,9
Totale	100,0	100,0	100,0
N. imprese	5.321	24.919	43.481

Nota: la classificazione per classe di addetti è desunta dall'archivio ASIA. Per le imprese non presenti in ASIA l'informazione non è disponibile (*missing*).
(1) Gli enti non commerciali sono esclusi da questa rappresentazione non essendo presenti in ASIA.
Fonte: elaborazione su dati dichiarazioni Agenzia delle entrate

distribuisce tra aliquota ordinaria e aliquote ridotte, con una quota non trascurabile riconducibile a queste ultime, soprattutto sul lato delle operazioni attive. Questo elemento segnala l'importanza delle agevolazioni IVA nel determinare il gettito complessivo e suggerisce come modifiche anche limitate alle aliquote o alle basi imponibili possano avere effetti rilevanti in termini di entrate e di distribuzione del carico fiscale.

La tabella 2 documenta la distribuzione delle imprese soggette alle tre imposte per classe di addetti, mettendo in evidenza la forte prevalenza di imprese di piccola dimensione, in particolare microimprese, in tutte e tre le basi imponibili considerate. Tale configurazione riflette la struttura produttiva locale (Tomelleri 2024) e ha implicazioni rilevanti per il disegno delle politiche fiscali, poiché suggerisce che anche interventi di carattere generale tendono a coinvolgere un numero elevato di imprese di dimensione ridotta. Allo stesso tempo, la presenza di una quota non trascurabile di osservazioni prive di informazione sulla dimensione aziendale segnala i limiti informativi derivanti dall'integrazione delle fonti amministrative, in particolare per i soggetti non coperti dall'archivio ASIA.

La tabella 3 riporta l'ammontare complessivo dell'imposta netta versata dalle imprese per classe di addetti, mettendo in evidenza una differenza tra la distribuzione del gettito e la distribuzione numerica delle imprese. Pur rappresentando la maggioranza dei soggetti, le microimprese contribuiscono a una quota relativamente contenuta del gettito complessivo, mentre le imprese di dimensione medio-grande, pur numericamente limitate, concentrano una parte rilevante delle entrate fiscali (cf. tabella 2). In termini assoluti, l'imposta netta complessivamente dichiarata nel 2019 ammonta a circa 258,3 milioni di euro per l'IRES (249,9 milioni

Tabella 3. Ammontare complessivo (in milioni di euro) dell'imposta netta per classe addetti. Anno 2019

Classe addetti	IRES	IRAP	IVA dovuta	IVA a credito
1-9 (micro)	64,0	34,7	363	131
10-49 (piccole)	61,1	18,4	241	110
50-249 (medie)	40,2	16,6	284	119
250+ (grandi)	25,3	7,4	68	55
Missing	59,3	18,7	151	55
Missing (ENC)	8,4	-	-	-
Totale	258,3	95,9	1.107	470

Fonte: elaborazione su dati dichiarazioni Agenzia delle entrate

di euro per le società di capitali e 8,4 per gli enti non commerciali) e 96 milioni per l'IRAP, mentre per l'IVA si osservano 1.107 milioni di euro di imposta a debito e 470 milioni di imposta a credito.

Questa asimmetria tra numerosità e contributo al gettito ha implicazioni importanti per la valutazione di possibili riforme fiscali, poiché suggerisce che interventi apparentemente neutri in termini di platea possono produrre effetti molto eterogenei in termini di entrate e di oneri concentrati su specifici segmenti del tessuto produttivo.

Ulteriori analisi mostrano una concentrazione del tessuto produttivo e quindi del versamento delle imposte nei comuni di Trento e Rovereto e nelle aree circostanti, riflettendo differenze e disparità tra le valli centrali e le zone montane. Infine, per quanto riguarda il settore di attività, notiamo una costante importanza in termini di imposta complessiva versata dal settore manifatturiero e commerciale per tutte le tre imposte, mentre, in termini di gettito, i servizi finanziari diventano particolarmente importanti per IRAP, e i trasporti per IRES e IVA.

Tutte queste informazioni sull'anagrafica d'impresa arricchiscono la comprensione del tessuto produttivo territoriale e forniscono punti di riferimento importanti per validare la corrispondenza tra l'economia territoriale e il dato da dichiarazione. Inoltre, queste caratteristiche d'impresa possono rappresentare ulteriori dimensioni di analisi negli studi di simulazione e d'impatto.

5. Simulazioni

L'obiettivo di questa sezione è illustrare le potenzialità di TREMOD-Imprese come strumento di simulazione ex ante di riforme fiscali. L'analisi si concentra in primo luogo su IRES e IRAP, per le quali la struttura delle dichiarazioni consente di ricostruire con elevato grado di precisione l'imposta netta. L'IVA, nella configurazione attuale di TREMOD-Imprese, è stata utilizzata principalmente per analisi descrittive in quanto le informazioni disponibili risultano ancora parziali. Nel box 1 si riportano tuttavia alcune potenziali linee di simulazione basate sulle informazioni disponibili, che potranno essere sviluppate in modo più sistematico in presenza di una base dati più dettagliata.

Prima di procedere all'implementazione degli scenari controfattuali, è necessario verificare che, a partire dalle componenti elementari osservabili nelle dichiarazioni, sia possibile ricostruire correttamente l'imposta netta effettivamente dichiarata. Per quanto riguarda l'IRES si osserva una quasi perfetta corrispondenza tra l'imposta simulata e l'imposta netta dichiarata; analogamente, per l'IRAP, nel confronto tra imposta netta da dichiarazione e imposta netta simulata, solo meno dello 0,1% dei casi non vengono simulati correttamente. Questa validazione garantisce che gli scenari controfattuali successivamente implementati si fondino su una rappresentazione coerente della normativa vigente[13].

Le simulazioni si basano sulla costruzione di scenari ipotetici alternativi[14]. Il confronto tra questi e la normativa vigente nel 2019 (*benchmark*) costituirà la stima d'impatto, ossia la variazione delle imposte versate causata dalla modifica. Oltre a conoscere l'effetto in termini di variazione di gettito, l'analisi consente di quantificarne l'eterogeneità e gli eventuali effetti redistributivi che ne conseguono.

È importante sottolineare che le stime così costruite sono di natura statica, cioè non incorporano reazioni comportamentali delle imprese (ad esempio, modifiche negli investimenti, nella localizzazione o nell'organizzazione produttiva). In altre parole,

13 Consideriamo simulati correttamente i casi con un margine di errore entro i 10 euro, che corrisponde allo 0,2% dell'imposta netta in media.

14 Si noti che le simulazioni vengono effettuate solamente sui casi che riusciamo a riprodurre correttamente.

Tabella 4. Scenari esemplificati. Versamento IRES reale, simulato e differenza, 2019 (milioni di euro)

Classe addetti	Versamento IRES 2019 reale	Scenario: eliminazione detrazioni		Scenario: eliminazione aliquote ridotte	
		Versamento IRES simulato	Impatto	Versamento IRES simulato	Impatto
1-9 (micro)	64,0	64,4	+0,4	64,0	+0,0
10-49 (piccole)	61,1	61,6	+0,5	61,1	+0,0
50-249 (medie)	40,2	40,4	+0,2	42,5	+2,3
250+ (grandi)	25,3	25,3	+0,0	25,3	+0,0
Missing (SC)	59,3	59,3	+0,0	59,3	+0,0
Totale SC	249,9	251	+1,1	252,2	+2,3
ENC	8,4	8,9	+0,5	11,3	+2,9
Totale SC e ENC	258,3	259,9	+1,6	263,5	+5,2

Fonte: elaborazione su dati dichiarazioni Agenzia delle entrate

Tabella 5. Scenari esemplificati. Gettito IRAP reale, simulato e differenza, 2019 (milioni di euro)

Classe addetti	Gettito IRAP 2019 reale	Scenario: aliquota standard da 2,68% a 3,9%		Scenario: eliminazione deduzioni regionali	
		Gettito IRAP simulato	Impatto	Gettito IRAP simulato	Impatto
1-9 (micro)	34,7	50,0	+15,3	34,9	+0,2
10-49 (piccole)	18,4	25,2	+6,8	18,6	+0,2
50-249 (medie)	16,6	20,6	+4,0	17,1	+0,5
250+ (grandi)	7,4	8,8	+1,4	7,7	+0,3
Missing	18,7	24,3	+5,6	18,8	+0,1
Totale	95,9	129,0	+33,1	97,0	+1,1

Fonte: elaborazione su dati dichiarazioni Agenzia delle entrate

assumendo che redditi, investimenti e altre variabili rimangano invariati, le simulazioni forniscono una misura dell'effetto contabile immediato a parità di condizioni.

A titolo illustrativo, vengono presentati due scenari di riforma dell'IRES: il primo riguardante il calcolo dell'imposta netta, il secondo le aliquote ridotte. Nel primo scenario si ipotizza l'eliminazione integrale delle detrazioni dall'imposta. In questo caso, il gettito IRES simulato ammonterebbe a 259,9 milioni di euro con un incremento di 1,6 milioni rispetto al *benchmark* (tabella 4). L'effetto complessivo risulta inferiore all'1% del gettito osservato, segnalando che, nel 2019, le detrazioni incidevano in misura relativamente contenuta sull'imposta netta aggregata. L'impatto contenuto della riforma appare inoltre distribuito in maniera piuttosto uniforme tra le diverse classi dimensionali d'impresa.

Nel secondo scenario si ipotizza l'eliminazione delle aliquote ridotte, mantenendo esclusivamente l'aliquota standard. In questo caso, l'impatto in termini di IRES versata sarebbe quantificabile in circa 5,2 milioni di euro, corrispondenti a un incremento di circa il 2% rispetto al gettito IRES 2019. L'effetto della riforma risulta inoltre eterogeneo per classe dimensionale: tra le società di capitali l'impatto si concentra sulle imprese di medie dimensioni, mentre per le altre classi l'effetto è pressoché nullo. L'impatto si concentra soprattutto sugli enti non commerciali (di cui, come abbiamo visto, non conosciamo la classe dimensionale), che beneficiano maggiormente di aliquote ridotte (cf. tabelle 1 e 3).

Anche per l'IRAP si possono effettuare simulazioni per valutare l'effetto del cambiamento di aliquote o l'effetto dell'eliminazione di deduzioni o detrazioni regionali. È importante considerare che data la natura locale dell'IRAP, oltre alle limitazioni comportamentali già menzionate sopra, queste simulazioni non tengono conto di eventuali cambiamenti comportamentali che possono portare alcuni soggetti a spostare le attività produttive in altri territori italiani.

A scopi dimostrativi di ciò che è possibile in termini di simulazioni IRAP, presentiamo due esempi che riguardano un cambiamento di aliquote e l'eliminazione delle deduzioni regionali (tabella 5). Il primo esempio simula l'effetto sul gettito complessivo di un cambiamento dell'aliquota standard dal 2,68% al 3,9% (aliquota *benchmark* italiana). Dato che la base imponibile non varia e che l'aliquota principale aumenta (quasi del 50%), non sorprende che il gettito aumenti del 34,5%, salendo approssimativamente

Box 1. Dimensioni normative utilizzabili per scenari controfattuali sulle tre imposte

IRES

Modifica base imponibile

- Abrogazione ACE (Aiuto alla Crescita Economica)
- Modifica rendimento nozionale ACE
- Introduzione soglie sui crediti di imposta
- Introduzione tetto a liberalità
- Introduzione tetto a legge 112/2016

Modifica aliquote

- Abolizione aliquota ridotta
- Modifica aliquota standard
- Aliquota differenziata in base al settore ATECO

Modifica imposta netta

- Modifica dell'ammontare complessivo delle detrazioni (es. abolizione totale, introduzione di un tetto)[15]

IRAP

Abolizione dell'imposta

- Per tutti o per alcuni soggetti in base alla forma giuridica (limitatamente alle informazioni da dichiarazione o da ASIA, laddove la forma giuridica è specificata)

Modifica aliquote

- Variazione dell'aliquota per specifici settori (limitatamente alle informazioni da dichiarazione o da ASIA, laddove il codice ATECO e/o il codice aliquota sono specificati)

Deduzioni e detrazioni[16]

- Rimozione
- Modifica (es. introduzione di un tetto)
- Introduzione di deduzioni per il costo del lavoro

IVA

Modifica base imponibile

- Basi imponibili attualmente esenti o non soggette all'imposta possono essere incluse nella base imponibile dell'aliquota ridotta o ordinaria.

Modifica aliquote[17]

- Aumento o riduzione dell'aliquota ordinaria
- Modifica delle aliquote ridotte principali (4%, 5%, 10%) o di quelle relative al settore agricolo.
- Sottoposizione delle acquisizioni intra-UE ad un'aliquota unica.

Nota: per tutte le imposte è possibile scomporre le simulazioni per varie caratteristiche d'impresa: classe di addetti, forma giuridica, settore, comunità di valle, per evidenziare le eterogeneità degli impatti.

15 L'attuale versione del modello considera esclusivamente l'ammontare aggregato delle deduzioni e delle detrazioni. Non essendo ancora implementata la stima delle singole componenti, allo stato attuale non è possibile valutare tramite simulazione modifiche riferite a specifiche tipologie di deduzioni o detrazioni.

16 Si veda nota 15.

17 Con i dati provenienti dalla dichiarazione IVA, non è possibile stimare cambiamenti di aliquote per beni o servizi specifici. Si potrebbe al contrario simulare una modifica delle aliquote sulla base del settore ATECO.

di 33 milioni in totale. L'impatto della riforma risulta eterogeneo per dimensione d'impresa e, in termini assoluti, l'aumento di gettito maggiore si osserva tra le microimprese, coerentemente con la struttura produttiva del territorio.

Nel secondo caso viene simulato l'effetto di un'eliminazione totale delle deduzioni regionali. In questo scenario gli introiti IRAP aumenterebbero di poco oltre il milione. L'aumento del gettito risulta inoltre distribuito in maniera piuttosto uniforme tra le diverse classi dimensionali d'impresa.

Il box 1 vuole sintetizzare, seppur in maniera non esaustiva, quali possono essere gli elementi su cui agire per la creazione di scenari ipotetici. Va precisato che non tutte le simulazioni indicate nel box sono immediatamente effettuabili con i dati sin qui acquisiti: in alcuni casi si rende necessario un allargamento della base dati. Ad esempio, se volessimo introdurre delle modifiche nel calcolo del reddito imponibile IRES sarebbe necessario disporre di tutti gli elementi che lo compongono[18]. Lo stesso vale per ipotetici scenari di modifica del reddito imponibile ai fini IRAP. Non avendo a disposizione le componenti dettagliate sul valore della produzione non possiamo risalire alla loro rilevanza ai fini della determinazione della base imponibile ed effettuare simulazioni a riguardo.

Inoltre, se volessimo stimare l'impatto dell'applicazione di un'aliquota ridotta dell'IRES sugli utili accantonati finalizzati a investimenti qualificati e incrementi occupazionali avremmo bisogno, ad esempio, di recuperare il quadro RU dei crediti cioè sapere se l'impresa ha sostenuto spese di ricerca e sviluppo, o spese di formazione per il personale su tecnologie avanzate o per investimenti in beni strumentali. Inoltre, non siamo in grado di fare simulazioni su specifiche detrazioni e deduzioni regionali poiché nei dati a disposizione queste due voci sono riportate come ammontare complessivo.

Conclusioni

Il presente contributo ha illustrato la costruzione e le prime applicazioni di TREMOD-Imprese, un modello basato sull'integrazione di dati fiscali amministrativi finalizzato alla valutazione ex ante della fiscalità d'impresa in un contesto sub-nazionale caratterizzato da autonomia finanziaria. L'analisi mostra come le dichiarazioni IRES, IRAP e IVA, opportunamente integrate con l'anagrafe delle imprese, possano costituire non solo una fonte informativa descrittiva, ma una base operativa per la simulazione di scenari normativi alternativi.

La rilevanza dello strumento risulta particolarmente evidente in relazione all'IRAP, il cui gettito affluisce direttamente al bilancio provinciale. In tale contesto, le simulazioni non rappresentano un esercizio meramente contabile, ma uno strumento potenzialmente utile alla programmazione delle entrate e alla valutazione preventiva degli effetti redistributivi delle scelte fiscali. Per IRES e IVA, l'analisi mantiene comunque una rilevanza per la Provincia autonoma di Trento in virtù dei meccanismi di compartecipazione al gettito. Essa consente di comprendere la struttura del gettito generato sul territorio e la distribuzione del carico fiscale tra tipologie di imprese, fornendo elementi informativi rilevanti per il *policy design*.

Quanto presentato rappresenta una prima fase di sviluppo del progetto. È attualmente in corso l'estensione temporale della base dati (2015-2023), con un ampliamento del dettaglio informativo relativo alle componenti delle basi imponibili, ai crediti d'imposta e alle deduzioni/detrazioni dichiarate. Tale arricchimento consentirà simulazioni più articolate e una maggiore aderenza alla complessità normativa. Un'estensione della copertura territoriale ad altre regioni va chiaramente oltre lo scopo di questo contributo e richiederebbe interazioni strutturate che non coinvolgano esclusivamente enti della Provincia di Trento. Tuttavia, l'inclusione dell'IVA nel modello Istat-MATIS (Caiumi e Di Biagio, 2016) potrebbe rappresentare il passo più prossimo all'esistenza di un equivalente di TREMOD-Imprese a livello nazionale.

18 L'imponibile IRES simulato a partire dall'utile (o dalla perdita) con l'aggiunta delle variazioni fiscali e al netto di interessi passivi non deducibili, liberalità, ACE e crediti d'imposta corrisponde alla base imponibile presente in dichiarazione in 9.000 casi mentre non corrisponde in 2.000. Occorre quindi verificare quali componenti mancano che vanno a concorrere alla costruzione della base imponibile.

Bibliografia

Azzolini D., Bazzoli M., De Poli S., Fiorio C., Poy S. (2017), Developing and validating regional microsimulation models: TREMOD: the tax-benefit model of the Italian Province of Trento, *Economia pubblica,* XLIV, n.1, pp.5-33

Bellucci C., Carta S., De Tollis S., Di Giacomo F., Manzo M., Bucci D., Curto D., De Grandis F., Sica F. (2023), *A Corporate Income Tax Microsimulation Model for Italy*, Working Paper 2023-17, Roma, MEF

Caiumi A., Di Biagio L. (2016), The Istat–MATIS corporate tax model, *Rivista di statistica ufficiale*, n.2, pp.131-159

Crato N., Paruolo P. (eds) (2019), *Data-Driven Policy Impact Evaluation. How Access to Microdata is Transforming Policy Design*, Cham, Springer Open

Figari F., Gandullia L., Leporatti L. (2019), I modelli di microsimulazione fiscale in ambito locale: il caso di LigurMOD1, *La finanza territoriale*, pp.171-195

Finke K., Heckemeyer J.H., Spengel C. (2010), *ZEW TaxCoMM–A Corporate Tax Microsimulation Model,* Mannheim, ZEW - Centre for European Economic Research

Gastaldi F. (2019), *MEDITA. A Corporate tax microsimulation model at PBO*, Torino, SIEP Annual Conference

Maitino M. L., Ravagli L., Sciclone N. (2017), Microreg: A Traditional Tax-Benefit Microsimulation Model Extended To Indirect Taxes And In Kind Transfers, *International Journal of Microsimulation*, 10, n.1, pp.5-38

Santoro F. (2003), *Federalismo fiscale e decentramento dell'IVA*, Documenti di lavoro dell'Ufficio Studi n. 7, Roma, Agenzia delle entrate

Shahnazarian H. (2011), A dynamic micro-econometric simulation model for firms, *International Journal of Microsimulation*, 4, n. 1, pp.2-20

Sutherland H., Figari F. (2013). Euromod the European Union tax-benefit microsimulation model, *International Journal of Microsimulation*, 6, n. 1, pp.4-26

Tomelleri A. (2024), Produttività e crescita economica, *Economia Trentina*, 73, n.3, pp.12-17

Martina Bazzoli
bazzoli@irvapp.it
Statistica sociale con esperienza nell'analisi dati e applicazione di tecniche econometriche per la valutazione di programmi e politiche pubbliche nel campo del mercato del lavoro, della formazione e delle politiche fiscali. Ha ottenuto la laurea specialistica in Scienze statistiche demografiche e sociali all'Università di Padova. È ricercatrice presso l'Istituto per la Ricerca valutativa sulle politiche pubbliche della Fondazione Bruno Kessler (FBK-IRVAPP) dal 2013. Tra le ultime pubblicazioni: (con Fiorio C., Di Carlo P., Figari F., Manzo M.), Self-employment income tax evasion and inequality, *Journal of Economic Behavior & Organization*, n.247, 2026.

Sonia Marzadro
marzadro@irvapp.it
Ricercatrice presso l'Istituto per la Ricerca valutativa sulle politiche pubbliche della Fondazione Bruno Kessler (FBK-IRVAPP). Si occupa di valutazione delle politiche pubbliche. I suoi interessi riguardano la disuguaglianza sociale, con particolare attenzione a mobilità sociale, povertà, impatti distributivi e opportunità educative, nonché all'uso delle tecnologie digitali nei processi formativi. Ha conseguito il dottorato in Sociologia e ricerca sociale presso l'Università di Trento. Tra le ultime pubblicazioni: (con Bazzoli N., Fiorio C., Pycroft J., Vergolini L.), Equal focus on inequality? approaches to distributional impact assessment in the national budget process across the EU, JCMS: *Journal of Common Market Studies*, n.6, 2024.

Annalisa Tassi
atassi@irvapp.it
Ricercatrice presso l'Istituto per la Ricerca valutativa sulle politiche pubbliche della Fondazione Bruno Kessler (FBK-IRVAPP). Ha ottenuto il dottorato in Economia presso l'università Friedrich-Alexander di Erlangen e Norimberga. I suoi interessi di ricerca riguardano l'economia pubblica, in particolare il comportamento caritatevole e frode ed evasione fiscale. Tra le ultime pubblicazioni: (con Bussy A.), Cross-border value-added tax fraud in the European Union, *Swiss Journal of Economics and Statistics*, n.1, 2025.

Carlo Fiorio
carlo.fiorio@unimi.it
Professore di Scienza delle finanze all'Università degli Studi di Milano e *Research Associate* di FBK-IRVAPP. Ha svolto gli studi a Pavia e alla *London School of Economics*. Si occupa di economia pubblica, con particolare attenzione all'analisi delle politiche fiscali e sociali, alla distribuzione dei redditi e alle disuguaglianze. Ha partecipato a numerosi progetti di valutazione delle politiche pubbliche, anche attraverso modelli di microsimulazione. Tra le ultime pubblicazioni: (con Chiappori P.A., Galichon A., Verzillo S.), Assortative matching on income, *Econometrica*, n.3, 2026.

Saggi

Ripensare la leadership: verso un nuovo modello di professionalità 5.0

Anna Grimaldi
INAPP

Alfonso Monsurrò
LUISS Università di Roma

La diffusione sempre più pervasiva dell'Intelligenza artificiale (IA) all'interno delle imprese implica un riposizionamento del modo in cui vengono prese le decisioni e disegnati i processi organizzativi. La domanda di ricerca che ha sorretto il lavoro che qui si presenta si chiede come l'introduzione dell'IA nei contesti produttivi ridisegna ruolo e competenze del leader. Il paper riporta i principali risultati ottenuti da un'indagine di campo e propone un modello di leadership 5.0 che traduce le evidenze empiriche in una struttura coerente di identità, aree d'azione e competenze.

Rethinking the leadership: towards a new professionalism 5.0 Model

The increasingly pervasive diffusion of Artificial intelligence (AI) within companies is leading to a repositioning of the way decisions are made and organizational processes are designed. The research question underlying the work presented here asks how the introduction of AI in production contexts reshapes the role and competencies of leaders. The paper reports the main findings from a field study and proposes a leadership model 5.0 that translates empirical evidence into a coherent framework of identity, areas of action, and skills.

DOI: 10.53223/Sinappsi_2026-02-12

Citazione
Grimaldi A., Monsurrò A. (2026), Ripensare la leadership: verso un nuovo modello di professionalità 5.0, *Sinappsi*, XVI, n.2, pp.174-188

Parole chiave
Innovazione delle imprese
Intelligenza artificiale
Leadership

Keywords
Innovation in enterprises
Artificial intelligence
Leadership

Introduzione

Nell'attuale orizzonte che si va delineando con crescente chiarezza, complessità e ampiezza sotto l'etichetta ormai consolidata di Industria 5.0, la diffusione sempre più pervasiva dell'Intelligenza artificiale all'interno delle imprese e dei sistemi produttivi non può essere interpretata come un mero aggiornamento tecnico o come una semplice evoluzione lineare delle tecnologie precedenti. Al contrario, essa implica un riposizionamento profondo, strutturale e radicale del modo in cui vengono prese le decisioni, disegnati i processi organizzativi, esercitata la governance e, più in generale, generato valore nelle strategie delle organizzazioni.

Il presente lavoro si colloca consapevolmente in questo tornante storico, assumendone tanto le tensioni e le ambiguità quanto le opportunità emergenti, e propone un percorso di ricerca che intreccia in modo sistematico un solido quadro teorico con un'articolata indagine sul campo. L'obiettivo è comprendere come la leadership, intesa come funzione e come pratica sociale, cambi forma, sostanza e modalità di esercizio quando i sistemi intelligenti diventano non più semplici strumenti di supporto, ma veri e propri partner operativi e cognitivi delle persone e delle organizzazioni.

L'ottica adottata è umanocentrica e organizzativa. Non si tratta, infatti, di celebrare l'automazione in chiave tecnicistica o deterministica, né di respingerla

per nostalgia di modelli passati ormai superati, ma piuttosto di chiarire in quali condizioni culturali, valoriali e di governance sia possibile trasformare l'Intelligenza artificiale in una leva di apprendimento collettivo, di qualità del lavoro e di benessere diffuso. Allo stesso modo, si vuole comprendere in quali contesti, invece, essa rischi di convertirsi in un moltiplicatore di opacità, sorveglianza e deresponsabilizzazione collettiva, compromettendo l'autonomia e la dignità professionale.

Su queste premesse si innesta l'oggetto di studio centrale: la leadership intesa come pratica situata, contestuale e relazionale, capace di orchestrare intelligenze complementari, umane e artificiali, di definire confini e responsabilità e di ricomporre in modo coerente il legame tra strategia e quotidiano, mantenendo in equilibrio efficienza e significato, performance e senso, obiettivi e valori.

Da questa prospettiva prende forma la domanda di ricerca[1] che attraversa l'intero lavoro e che culmina nel modello propositivo che qui si presenta per attivare una riflessione scientifica sull'argomento: l'introduzione dell'Intelligenza artificiale negli attuali contesti produttivi e organizzativi ridisegna ruoli, aree d'azione e competenze della leadership? E quali premesse culturali, valoriali e assetti organizzativi rendono possibile un'adozione dell'IA che non sostituisca l'uomo, ma ne amplifichi la capacità d'azione, salvaguardandone al contempo la centralità, l'autonomia e la dignità professionale?

Il lavoro non assume scorciatoie interpretative, né si rifugia nelle retoriche polarizzanti del "tutto è cambiato" o del "non cambia nulla". Esplora, invece, gli assi concettuali che emergono dall'esperienza concreta di manager e leader impegnati[2] nei percorsi di integrazione dell'IA nelle loro strutture. Queste esperienze vengono rilette alla luce delle principali tradizioni teoriche sulla leadership e dei riferimenti più pertinenti del paradigma 5.0, che pone al centro la collaborazione uomo–macchina in chiave etica e sostenibile.

L'ipotesi di fondo che orienta il lavoro è che l'Intelligenza artificiale agisca simultaneamente come moltiplicatore di possibilità e come fattore di pressione sulle identità professionali e sulle architetture decisionali delle organizzazioni. Ne consegue che sarà la qualità della leadership, radicata in culture coerenti, in sistemi di governance trasparenti e in visioni condivise, a determinarne gli esiti effettivi e la reale portata trasformativa.

Senza voler qui fare una disamina sulle varie teorie e dei tanti modelli di leadership che si sono succeduti nel corso del tempo (Avolio e Bass 1988; Chamakiotis *et al.* 2021; Di Fabio e Gori 2021, Di Fabio *et al.* 2023; Fiedler 1964; Goleman 1998; Hersey e Blanchard 1988; House e Mitchell 1977; Liu *et al.* 2020; López-Figueroa *et al.* 2025; Stogdill 1974; Yukl 2013) e che hanno visto il progressivo passaggio da una concezione del leader come eroe carismatico, quasi una peculiarità personale del singolo individuo, ad una concezione multifattoriale che vede la leadership come processo dinamico e collettivo, risultato dell'interazione tra persone, tecnologie, ruoli e contesti, si vuole invece porre enfasi su un paradigma che passa da un pensiero centrato sul controllo e sull'autorità a un modello orientato alla collaborazione, alla fiducia e alla trasparenza e che quindi svetta verso una visone della leadership come capacità di interpretare la complessità e di dare coerenza a sistemi plurali, dimensione questa particolarmente rilevante nell'epoca attuale. In tale concezione un leader non deve solo essere in grado di gestire le persone, ma deve definire e soprattutto trasmettere valori in un'organizzazione. Già Edgar Schein (1985) nel definire la cultura aziendale come un insieme di valori e credenze condivisi sottolineava quanto fosse importante l'esempio del leader: la prontezza e le strategie con cui un leader gestisce i momenti di crisi, cosa valorizza e come si relaziona con i propri collaboratori, definiscono quell'insieme di valori condivisi. I nuovi leader per avviare e sostenere il cambiamento, competenze oggi primarie, devono diventare "ricercatori di informazioni" (Schein, 2010) per immergersi nella cultura e nei valori del gruppo e costruire così relazioni fondate sulla fiducia e sulla stima reciproca.

Allo stesso modo non si vuole qui tracciare la storia e l'evoluzione delle innumerevoli visioni dell'Intelligenza artificiale e delle sue conseguenze sul futuro del lavoro (Alshaibani *et al.* 2025; Burgess 2024; De Cremer 2022; Laybats e Tredinnick 2024; Machado e Davim 2023; Oh e Ki 2024; Rezazadeh *et al.* 2025) ma si assume una prospettiva dell'IA come ecosistema di interazioni, di mediazione cognitiva che modifica i modi di conoscere, di interpretare la realtà e di agire all'interno delle organizzazioni dove la co-decisione uomo–macchina diventa un tema centrale e cruciale. Tale concezione deve fondarsi su principi di inclusione, *accountability* e tutela della dignità del lavoro; pertanto, in tale contesto la leadership deve progettare, presidiare e sostenere i contesti organizzativi così che l'IA amplifichi l'intelligenza sociale e collettiva, anziché sostituirla o disarticolarla.

Partendo da tali premesse concettuali è stata concepita un'indagine di campo che ha visto la

1 La ricerca è stata realizzata in occasione di un lavoro di tesi magistrale alla Luiss Università di Roma.

2 Si coglie l'occasione per ringraziare tutti gli opinion leader che hanno contribuito alla realizzazione di questo lavoro.

testimonianza di autorevoli *opinion leader* impegnati in processi di riorganizzazione nella convinzione che la testimonianza di chi sperimenta quotidianamente i processi di cambiamento ne possa evidenziare opportunità e criticità.

1. Metodologia

Il disegno di ricerca ha privilegiato un approccio qualitativo ritenuto il più idoneo per indagare in profondità il fenomeno analizzato e per coglierne la complessità nelle sue molteplici dimensioni. La domanda di ricerca, che, come anticipato, riguarda l'evoluzione del ruolo e delle competenze della leadership nell'Industria 5.0 in relazione all'introduzione dell'Intelligenza artificiale, richiede infatti di andare oltre l'osservazione delle azioni messe in atto dai leader, per comprendere come essi sollecitino, interpretino, e soprattutto quale significato attribuiscano a tali trasformazioni. Nello specifico è stato adottato il paradigma della *self-reflectivity*, intesa come l'assunzione da parte del ricercatore di uno sguardo consapevole e riflessivo, capace di sospendere giudizi personali e interpretazioni precostituite per accogliere pienamente la prospettiva dell'intervistato. Ciò ha significato prestare attenzione non soltanto al contenuto verbale delle risposte, ma anche agli aspetti relazionali e comunicativi che le accompagnavano: l'intervista è stata concepita non come una mera tecnica di raccolta di dati, bensì come un processo di co-costruzione della conoscenza, in cui intervistatore e intervistato contribuiscono insieme a delineare il senso di ciò che emerge (Ancora *et al.* 2023). La *self-reflectivity* ha anche orientato la fase di analisi, dove i dati sono stati riletti nella consapevolezza che ogni affermazione va interpretata in relazione al ruolo, al contesto e al vissuto di chi la pronuncia. In questo modo è stato possibile ridurre il rischio di proiezioni indebite per restituire un quadro il più possibile fedele e autentico delle esperienze narrate dai partecipanti. È stato così attribuito valore alla dimensione soggettiva della narrazione, riconoscendo che la comprensione di un fenomeno articolato, come quello oggetto di indagine, passa anche attraverso le storie e le interpretazioni di chi lo sperimenta quotidianamente. In aggiunta, la natura dialogica dell'intervista ha reso possibile seguire il flusso spontaneo della conversazione, aprendo la strada ad approfondimenti non previsti ma rilevatisi di particolare interesse per la ricerca e per l'individuazione di piste di riflessione e di sviluppi futuri. L'orientamento qualitativo, in questa prospettiva, non si è limitato a documentare comportamenti e opinioni, ma ha cercato di esplorarne le motivazioni profonde, le radici interpretative e il significato attribuito dagli intervistati alla propria esperienza di leadership in un contesto caratterizzato da rapida trasformazione tecnologica. È stato adottato il format dell'intervista semi-strutturata, a garanzia sia di un orientamento comune che assicurasse comparabilità dei dati sia della specificità di ogni interazione. In questo modo, ogni intervista ha conservato un nucleo tematico condiviso senza rinunciare alla ricchezza narrativa e alla specificità dei singoli contributi. Ne è derivato un corpus di dati qualitativi capace di restituire un quadro ampio e articolato del fenomeno, mettendo in luce punti di convergenza tra le esperienze raccolte, ma anche differenze riconducibili ai contesti organizzativi, ai settori di appartenenza e ai percorsi professionali individuali. Il protocollo di intervista è stato elaborato con l'obiettivo di assicurare coerenza all'intera fase di raccolta dei dati, bilanciando l'esigenza di rigore metodologico con la necessità di garantire flessibilità esplorativa. La sua definizione è partita dalla domanda di ricerca, traducendo in quesiti operativi le principali aree tematiche individuate: l'impatto dell'Intelligenza artificiale sui mercati e sulle organizzazioni, il ruolo della leadership nei processi di cambiamento, le competenze richieste ai leader nell'Industria 5.0, il peso della cultura e dei valori aziendali e le sfide emergenti in un contesto tecnologico in rapida trasformazione. La durata media dei colloqui è stata di circa quarantacinque minuti. Tutte le interviste sono state registrate, previa autorizzazione degli intervistati, così da poter procedere con una trascrizione integrale e accurata dei contenuti.

Gli intervistati sono stati individuati mediante un campionamento intenzionale, finalizzato a includere figure in grado di apportare contributi significativi e tra loro diversificati rispetto alla domanda di ricerca. Il criterio principale di selezione ha riguardato professionisti che ricoprissero ruoli di responsabilità in contesti organizzativi già coinvolti in processi di trasformazione digitale o caratterizzati da un'interazione rilevante con l'Intelligenza artificiale. Il campione raccoglie profili con consolidata esperienza nella gestione di persone e di processi complessi, responsabili di aree strategiche impegnate nell'introduzione di tecnologie emergenti e leader di funzioni aziendali particolarmente esposte agli effetti dell'innovazione. Sono stati quindi coinvolti HR manager nella convinzione che tali elementi li rendono osservatori privilegiati per comprendere l'evoluzione del ruolo del leader in scenari caratterizzati da cambiamenti tecnologici e professionisti con responsabilità in ambiti legati all'innovazione tecnologica e alla trasformazione digitale, capaci di offrire un contributo diretto su come l'IA venga effettivamente

integrata nei processi organizzativi e su quali siano le relative implicazioni strategiche e operative. In alcuni casi, la scelta è ricaduta su partecipanti in grado di unire una visione manageriale di ampio respiro con un'esperienza concreta nell'utilizzo delle nuove tecnologie, così da fornire un punto di vista bilanciato tra strategia e operatività[3]. La composizione eterogenea ha garantito un equilibrio tra chi sperimenta quotidianamente le sfide della leadership e chi, pur provenendo da contesti più tecnici, è in grado di evidenziarne le conseguenze in termini di cultura aziendale, processi organizzativi e dinamiche relazionali. Questa varietà di prospettive è stata ritenuta un elemento di valore, poiché ha consentito di integrare dimensioni manageriali, tecniche e strategiche, restituendo una lettura articolata e multidimensionale del fenomeno. L'inclusione di partecipanti appartenenti a settori differenti, pur mantenendo coerenza con l'oggetto della ricerca, ha permesso di mettere a confronto esperienze eterogenee, facendo emergere sia punti di convergenza sia differenze nell'approccio alla leadership e all'adozione dell'Intelligenza artificiale.

Per tutelare l'anonimato dei partecipanti, pur mantenendo la possibilità di collegare ogni dichiarazione al contesto professionale di provenienza, è stato adottato un sistema di codifica univoco. Questo approccio ha permesso di garantire la tracciabilità delle informazioni, rispettando al contempo l'impegno di riservatezza definito nel consenso informato. L'impiego dei codici ha inoltre semplificato la fase di analisi comparativa, consentendo di raggruppare e confrontare le affermazioni in base a caratteristiche quali settore di appartenenza, livello di responsabilità o area di competenza.

Al termine della fase di raccolta, la trasposizione è stata effettuata secondo un criterio "quasi verbatim", per passaggi interpretativamente complessi sono state annotate osservazioni, mantenendo separate queste note dal testo principale per preservarne l'integrità. Una volta completata e validata la fase di trascrizione, il corpus è stato sottoposto a un processo di codifica tematica con approccio induttivo, ossia lasciando che le categorie emergessero direttamente dai dati piuttosto che applicare griglie precostituite. La segmentazione del testo in unità di significato è avvenuta in maniera rigorosa, a ciascuna unità è stato associato un codice descrittivo, sintetico ma sufficientemente chiaro da consentire un'interpretazione univoca nella fase successiva di aggregazione. Il sistema di codifica ha mantenuto costantemente il riferimento al codice dell'intervistato, così da garantire la possibilità di risalire al contesto specifico in cui l'affermazione era stata formulata. In parallelo, sono state inserite note analitiche per registrare impressioni, connessioni con altri punti del corpus e possibili piste di approfondimento. Queste note non sono state trattate come dati, ma come strumenti di supporto al ragionamento interpretativo. Completata la codifica tematica, l'analisi si è concentrata sul confronto trasversale delle unità di significato per identificare ricorrenze, connessioni e differenze tra le affermazioni. L'approccio adottato è stato esplorativo: non si è cercata conferma di schemi predefiniti, ma si è lasciato che le relazioni tra i codici delineassero la mappa tematica dello studio. I pattern sono stati individuati considerando non solo la frequenza con cui determinati concetti comparivano nelle interviste, ma anche la profondità e la varietà delle argomentazioni a essi associate. In alcuni casi, la rilevanza di un tema non derivava dal numero di occorrenze, bensì dalla sua importanza strategica rispetto alla domanda di ricerca o dalla coerenza con altre evidenze emerse.

2. Principali evidenze empiriche

L'analisi del contenuto delle interviste ha permesso di fotografare una serie di temi chiave che descrivono come la leadership stia evolvendo nell'era dell'Industria 5.0 e dell'Intelligenza artificiale. I contributi dei testimoni mostrano una prospettiva articolata che spazia dagli effetti dell'IA sui mercati e sui modelli di business, alle modalità con cui i leader gestiscono il cambiamento e sviluppano competenze sempre più complesse. Viene inoltre messo in luce il ruolo cruciale della cultura organizzativa e dei valori aziendali, così come le sfide che accompagnano l'integrazione delle nuove tecnologie. I risultati offrono quindi una visione d'insieme che intreccia dimensione tecnologica, organizzativa e umana, restituendo un quadro ricco e multidimensionale. Emergono cinque fondamentali nuclei tematici di particolare rilievo.

Il primo riguarda l'ambivalenza dell'IA, percepita simultaneamente come acceleratore di performance e possibile fattore di frattura. Sul versante

3 Sono stati intervistati 10 opinion leader con ruoli apicali di governance di 10 diversi enti: Co-fondatore di società di consulenza e formazione manageriale, Chief Operating Officer di Fondazione di ricerca non profit, Presidente CdA di Azienda di formazione digitale, Head of Technology Innovation di Azienda del settore fashion & luxury Brunello Cucinelli, Dirigente responsabile PNRR del Ministero della Cultura, HR Director di Società di consulenza multinazionale, HR Director di Azienda del settore fashion & luxury, Senior Manager HR Intelligence di Azienda di soluzioni analytics e software, Senior Research Director di Società di ricerche di mercato, Managing Director di Società di consulenza strategica.

dell'efficienza, i manager intervistati descrivono cicli decisionali più rapidi grazie alla disponibilità di dati affidabili, a simulazioni di scenario più accurate, a una migliore anticipazione delle criticità e a una pianificazione più precisa e personalizzata. Tuttavia, le stesse capacità diventano fragilità se innestate in culture organizzative incoerenti o in processi poco trasparenti: la velocità può facilmente scivolare nella fretta, l'algoritmo può essere usato retoricamente per giustificare decisioni non discusse, e il linguaggio tecnico può erigersi a barriera invece di abilitare comprensione e dialogo. In breve, l'IA amplifica ciò che trova: dove prevalgono chiarezza di scopi, tracciabilità e confronto aperto, tende a potenziare l'azione; dove invece dominano controllo difensivo, sfiducia o retorica tecnocratica, irrigidisce tensioni e paure.

Il secondo nucleo tematico mette al centro la cultura e i valori come determinanti fondamentali della riuscita. Gli intervistati convergono nel ritenere che la tecnologia "assorba" e rifletta il clima in cui viene introdotta. In ambienti improntati a trasparenza, responsabilizzazione diffusa e apprendimento continuo, l'Intelligenza artificiale viene percepita come una leva di autonomia, innovazione e miglioramento; al contrario, in organizzazioni dominate dal controllo, dall'ambiguità tra valori dichiarati e pratiche reali, essa genera resistenze e timori. Il leader, in questo scenario, funge da mediatore tra strumenti e significati: stabilisce confini e responsabilità, istituisce spazi di confronto, legittima l'errore come occasione di apprendimento e promuove un linguaggio condiviso su temi come spiegabilità, auditabilità, soglie di confidenza, inclusione ed equità.

Il terzo nucleo riguarda il governo del cambiamento. L'adozione dell'IA non è un semplice passaggio tecnico, ma una trasformazione che tocca identità, ruoli e abitudini. Si rivela efficace una leadership capace di combinare chiarezza di visione e gradualità di attuazione, rigore operativo e attenzione alle dimensioni emotive, comunicazione franca di opportunità e rischi, e legittimazione del dubbio come risorsa critica. Le resistenze, se ascoltate e integrate, diventano prudenza operativa e fonte di apprendimento; se ignorate, si traducono in disimpegno o opposizione silente. Ricorrente è la preoccupazione per il *lock-in* tecnico o cognitivo: il leader efficace preserva capacità di valutazione interna, promuove standard aperti e interoperabilità, evita dipendenze da singoli fornitori e difende spazi deliberativi umani per le decisioni più sensibili.

Il quarto nucleo riguarda le competenze: la tradizionale dicotomia tra *hard* e *soft skills* non basta più. Qualità decisionale in contesti "aumentati" significa unire alfabetizzazione ai dati, consapevolezza dei limiti inferenziali, capacità di audit e tracciabilità, ma anche empatia, ascolto, negoziazione e chiarezza comunicativa nell'incertezza. A queste si affianca la competenza di contesto, la cosiddetta "visione ad elicottero", che consente di cogliere segnali deboli, mappare interdipendenze, collocare scelte locali entro visioni sistemiche e decidere in presenza di trade-off espliciti. Il confine uomo–macchina diventa così un oggetto di design deliberato, non una linea invalicabile.

Il quinto nucleo tematico ha a che fare con le sfide. Le interviste hanno evidenziato diverse sfide che i leader affrontano nell'integrare l'Intelligenza artificiale all'interno dei processi organizzativi. Tali sfide non si limitano agli aspetti tecnici o infrastrutturali, ma coinvolgono anche dimensioni culturali, etiche e relazionali, determinando la reale sostenibilità del cambiamento. Un primo nodo critico riguarda la gestione dei rischi legati all'IA, come i bias algoritmici, la scarsa trasparenza di alcune decisioni automatizzate e la protezione dei dati sensibili. I partecipanti hanno sottolineato che, a prescindere dal livello di autonomia dei sistemi, la responsabilità finale resta sempre nelle mani delle persone. Ciò richiede un controllo attivo dei processi e una supervisione costante, evitando di affidare in modo acritico decisioni strategiche agli strumenti tecnologici. Un'altra sfida riguarda la gestione della complessità e dell'incertezza. La rapidità con cui le tecnologie si evolvono rende difficile pianificare a lungo termine, obbligando i leader a prendere decisioni con dati incompleti e scenari mutevoli. In questo contesto, diventano fondamentali capacità di sintesi, flessibilità e apertura a rivedere le strategie in base alle nuove evidenze. Le interviste hanno anche evidenziato l'importanza di bilanciare tecnologia e componente umana. L'IA può aumentare l'efficienza e supportare le decisioni, ma se non gestita con attenzione rischia di indebolire le relazioni interpersonali. I leader devono assicurare che l'adozione tecnologica rafforzi, anziché compromettere, coesione, fiducia e motivazione dei team.

3. Verso un nuovo modello di leadership

La leadership, come abbiamo visto, è una delle aree più studiate del comportamento organizzativo e assume sempre più rilevanza in un momento in cui i processi produttivi stanno modificandosi radicalmente grazie anche all'avvento dell'Intelligenza artificiale. Ma se l'accumulo dei dati empirici e la molteplicità dei costrutti costituisce senza dubbio una ricchezza scientifica e operativa, ad oggi non si può dire di avere una visione e una conoscenza univoca e integrata del fenomeno ma anzi è in corso una riflessione culturale sulla

concezione della leadership che porta scienziati e ricercatori a ipotizzare definizioni e modelli di professionalità che di volta in volta pongono enfasi, a seconda della disciplina di appartenenza del ricercatore, talvolta a variabili più soggettive e individuali talvolta a variabili più organizzative e contestuali. Alla luce delle tante evidenze empiriche tracciate e rappresentate dai testimoni privilegiati che sono stati intervistati nella fase di campo di questo lavoro si può asserire che la leadership è sempre più un ruolo complesso di influenzamento degli altri. In altri termini la leadership corrisponde all'uso di un'influenza non coercitiva per dirigere e coordinare le attività dei membri di un gruppo organizzato verso il raggiungimento degli obiettivi del gruppo stesso (Jago 1982). Questa non banale operazione implica diverse funzioni (Wagner e Hollenbeck 1992):

- generare e mantenere il livello di impegno e di tensione richiesto ai singoli individui che compongono un gruppo
- direzionare lo sforzo del gruppo lungo prospettive che promuovono la sopravvivenza del gruppo ed il raggiungimento degli obiettivi
- gestire i compiti del gruppo e le dinamiche relazionali
- facilitare e mantenere l'appartenenza al gruppo, tenendo uniti gli individui, centrandoli sul compito e sugli obiettivi e soddisfacendo i bisogni dei membri del gruppo.

Le qualità e le competenze di un leader che sempre più acquisiscono valenza multiprofessionale e multisettoriali e si collocano nell'intreccio tra hard e soft skill, hanno tre polarizzazioni diverse ma tra di loro interattive.

La prima valenza è verso sé stesso: intendendo con questo la fiducia nelle proprie qualità, l'autostima e la *self confidence* ma anche motivazione autorealizzativa, sviluppo dell'identità propria (e del proprio gruppo/organizzazione), autocontrollo emotivo e tolleranza delle ansie di rischio e sociali, coerenza valoriale evitando distanze tra "dire e fare". E ancora, spiccata autocritica per gestire il proprio cambiamento, i propri errori e le proprie scelte.

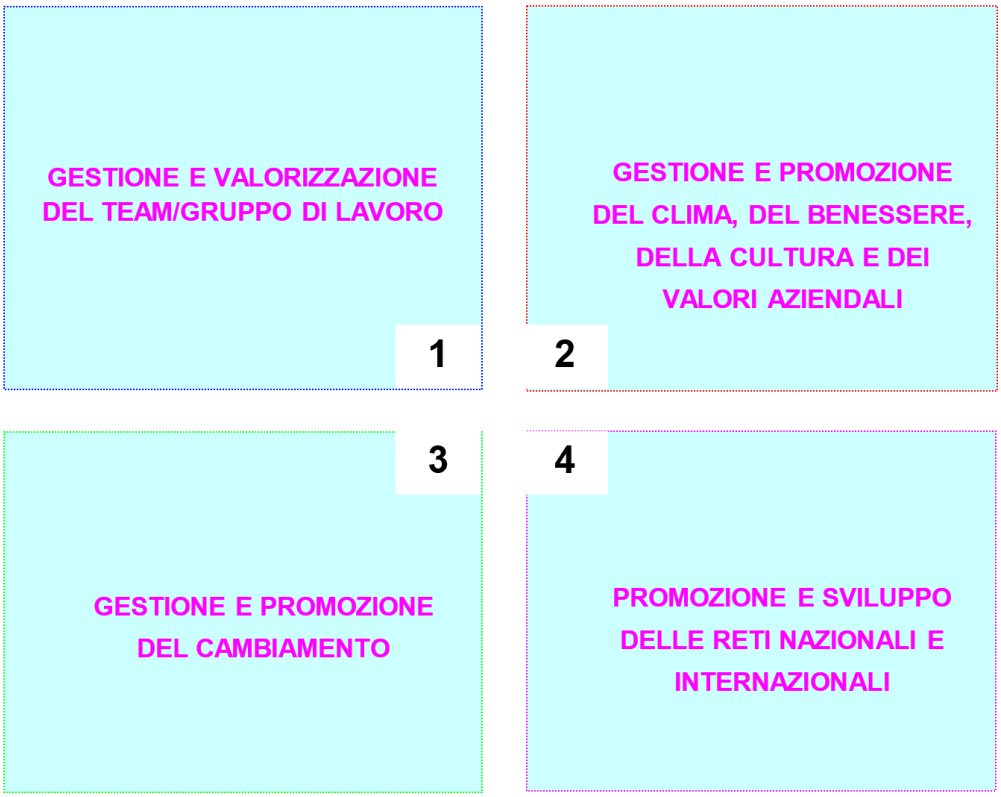
La seconda valenza è verso gli altri: un buon leader deve essere carismatico per ottenere rispetto senza pretenderlo, deve essere sfidante e supportivo per sostenere il gruppo nelle difficoltà, sollecitare motivazione e animare energie. Deve essere un ascoltatore attivo ed empatico per raccogliere idee, pareri, opinioni, e comprendere gli altri nelle loro esigenze più profonde. È importante anche che sia un buon gestore dei conflitti, cercando di sciogliere gli inevitabili nodi e di dirompere le divergenze collettive. E sempre di più deve essere non solo un facilitatore ma anche e soprattutto un promotore del cambiamento.

La terza valenza è verso l'ambiente perché ogni azienda/ente/istituto è collocato in un contesto in cui e con cui deve interagire e integrarsi. Deve essere quindi un conoscitore profondo della cultura organizzativa, delle reti territoriali, intuitivo e innovativo deve saper cogliere e sviluppare segnali deboli da tradurre in mete sfidanti, anche in condizioni di rischio e di incertezza. Deve ideare, anticipare e sperimentare innovazione e cambiamento. In altre parole, deve essere portatore di "visione ad elicottero": visione globale sfidante orientata al medio termine, verso la quale guida i collaboratori.

In considerazione di questa breve sintesi dei principali risultati emersi da questo lavoro si vuole qui provare a delineare un modello di professionalità per un leader ai tempi dell'industri 5.0, ipotizzando aree di intervento e di attività principali, nel tentativo di costituire un punto integrato di partenza per i successivi sviluppi riflessivi e concettuali per ripensare la leadership e pervenire ad una innovativa concezione di tale ruolo professionale. Se si condivide l'idea che questa figura va pensata come un "gestore" in grado di progettare e promuovere dispositivi e strumenti innovativi a supporto dello sviluppo e della qualificazione dei gruppi ma anche delle organizzazioni di appartenenza (e non solo), allora le funzioni da svolgere sono complesse e comportano un profilo di competenze articolato. In altri termini, tali funzioni di gestione richiedono di saper mantenere i rapporti con tutto il personale dell'ente/azienda ma anche con le altre imprese e le istituzioni del sistema, fornendo consulenza e supervisione nella definizione di strategie regionali, nazionali e internazionali così da costituire un punto di riferimento per tutte le attività e i progetti attuali e futuri. Il leader deve, pertanto, disporre di competenze articolate nella gestione del proprio ruolo, nella conduzione e nell'analisi del lavoro, nell'organizzazione dei piani e dei progetti di intervento, nella gestione/implementazione dell'assicurazione/controllo di qualità, nella realizzazione di disegni strategici d'impresa innovativi.

In figura 1 vengono visualizzati gli ambiti di attività nella loro globalità, senza attribuire a nessuno di essi un ordine gerarchico di importanza. Anzi, si potrebbe ipotizzare che ciascuno di questi ambiti corrisponda a una funzione professionale specifica che non in tutti i contesti organizzativi può essere ugualmente declinata. In altre parole, mentre il profilo professionale è naturalmente unico e le competenze assimilabili, le funzioni possono variare di intensità e declinate in attività specifiche diverse a

Figura 1. Principali aree di attività di un leader



Fonte: elaborazione degli Autori

seconda delle imprese/sistemi/ strutture operative in cui si colloca.

Relativamente alla gestione e alla valorizzazione del team/gruppo di lavoro, l'evoluzione organizzativa degli ultimi anni ha posto sempre più enfasi *al gruppo di lavoro e al lavoro di gruppo* e all'importanza di implementare e sviluppare una *mentalità collaborativa e inclusiva* nelle imprese che potenzi l'efficacia di un gruppo e renda possibili performance di successo. Ma quando si parla di gruppo di lavoro, soprattutto se in esso sono rappresentate competenze e funzioni multidisciplinari, non possiamo non considerare alcune dimensioni che lo declinano e lo caratterizzano in maniera molto più determinante di quanto le stesse dimensioni non si riferiscano ad ogni singola figura professionale e che, allo stesso tempo, costituiscono delle sfide da percorrere per prefigurarsi scenari innovativi ma anche dei vincoli che ne possono ostacolare il buon funzionamento. Vogliamo qui tracciare quelle dimensioni che, è convinzione di chi scrive, rappresentano gli elementi principali su cui fondare un valido lavoro di équipe (D'Emilione *et al.* 2020). La prima fa riferimento al senso di appartenenza. Sentire di far parte di un gruppo che condivide obiettivi, altrimenti preclusi alla singola professionalità, che si assume rischi perché sa di poter disporre del sostegno di tutti i membri del gruppo, ha indubbiamente anche degli effetti positivi sula resilienza rispetto all'impatto dei fattori stressanti, rafforza le capacità di *coping* e il senso di autoefficacia, cosa che comporta anche un aumento generale dell'autostima. Le singole identità professionali si fondono e si rafforzano in un'identità complessiva di gruppo che fa riferimento all'insieme delle risorse percepite come disponibili a offrire supporto di tipo emozionale, cognitivo e operativo. In tale ottica l'identità è plurima perché in costante rapporto con gli altri e dinamica perché in continuo adattamento e rappresenta un processo di sviluppo sotteso all'agire collettivo, in quanto membro di una comunità professionale in cui vengono condivise modalità di azione, di lettura e di analisi del contesto sociale e organizzativo. Se diversamente tale senso di appartenenza è carente, anche solo in parte, e ciò spesso accade per gelosie e stereotipi legati alle singole professionalità, si corre il rischio di chiudersi in una sorta di "solitudine professionale" con conseguenze negative sulla motivazione, sull'impegno nelle difficoltà, sui livelli di aspirazione e sul fronteggiamento delle frustrazioni e dello stress.

Una seconda dimensione su cui fondare un significativo lavoro di équipe fa riferimento alla corresponsabilità intesa come processo di assunzione delle decisioni che chiama in causa la natura coscienziale di ogni singola professionalità in relazione con tutte le altre ma anche con le organizzazioni e le istituzioni di appartenenza, in un continuum di dialogo e di confronto. Il processo deve sempre tener conto dei rischi e dei benefici che ne possono derivare. E deve fare riferimento a tecniche di *decision making* dove siano ben definite le interazioni tra singole responsabilità, procedure gestionali e organizzative nonché strumenti di controllo. La mancanza della corresponsabilità di un'équipe ha effetti negativi sulla performance della stessa fino ad arrivare alla piena disfunzionalità del servizio.

Una terza dimensione fa riferimento all'integrazione intesa anche come capacità di tenere insieme la propria diversità professionale con quella altrui. La concezione di diversità professionale ha un'accezione contenutistica, che si riferisce agli specifici compiti assegnati ai singoli profili, ma anche un'accezione metodologica riferendosi in questo caso alle peculiarità delle procedure che i contesti organizzativi attribuiscono alle diverse professioni. Una reale integrazione, che non significa omologazione né va confusa con la perdita di autonomia, è un processo di apprendimento continuo che impara a riconoscere e valorizzare le differenze a favore di una mappa concettuale e operativa di obiettivi, compiti e procedure, dove siano esplicitate integrazioni ma anche e confini e specificità professionali. La mancanza di integrazione porta alla separazione di compiti e attività e non assicura continuità al servizio.

Correlata alle dimensioni precedentemente presentate è la condivisione ma ci piace qui distinguerla dall'integrazione in quanto intesa come la capacità di riconoscersi nella specificità del gruppo di appartenenza con la consapevolezza che l'équipe, in quanto insieme di diverse professionalità, possiede un fervore concettuale e realizzativo impensabile per la singola persona. Seguendo un approccio psicosociale, l'équipe è qualcosa di diverso dalla somma delle parti, e solo se c'è vera condivisione, la pluralità, l'interazione e il legame professionale che la caratterizzano diventano motivo e propulsione al cambiamento. E infine come ultima ma non meno importante dimensione c'è il tema delle competenze, tematica molto cara a chi scrive, soprattutto se riferita ai modelli di professionalità e declinata nell'intreccio tra aree di attività e funzioni organizzative (Grimaldi 2009; 2010; 2025). Non vogliamo qui fare una dissertazione sul concetto di competenza e sulla sua definizione ben argomentata dalla vasta letteratura in materia ma vogliamo sottolineare l'importanza di pervenire a chiari ed espliciti modelli in riferimento a specifiche professionalità e questo è tanto più vero là dove si parla di équipe multiprofessionali. Modelli quindi che devono porre enfasi all'insieme delle caratteristiche legate alle singole posizioni organizzative e di lavoro, all'insieme delle caratteristiche professionali, al potenziale di qualità che può essere trasferito da contesti diversi, ai vincoli e alle opportunità dei soggetti organizzativi di riferimento nonché al confronto sociale e alle aspettative e agli obiettivi di policy. Se si condivide quanto fin qui argomentato allora non si può non considerare il difficile compito di un leader che deve innanzitutto conoscere il suo team di lavoro nelle aspettative, nelle esigenze, nelle competenze di ogni singolo componente e poi con empatia e senso di responsabilità deve motivare, coordinare, delegare, dare feedback efficaci, gestire i conflitti per poter pervenire ad un vero, efficace ed efficiente lavoro di gruppo.

Relativamente alla seconda area di intervento "Gestione e promozione del clima, del benessere, della cultura e dei valori aziendali" va sottolineato che questa è una dimensione in continua evoluzione ed è molto incisiva nel determinare un clima positivo favorevole allo sviluppo di una buona e sostanziale produttività. Il clima e il benessere organizzativo sono dimensioni rilevanti che determinano senso di appartenenza alla cultura organizzativa e influiscono sulla soddisfazione e sulla produttività. Il benessere organizzativo può essere inteso come la capacità di un'organizzazione di promuovere e mantenere un alto livello di salute fisica, psicologica e sociale dei lavoratori. In letteratura c'è un ampio accordo nel ritenere che lo stato di benessere di un'organizzazione derivi da un insieme multiplo di dimensioni, tra le quali il clima organizzativo, inteso come l'atmosfera prevalente che si respira, costituisce certamente un fattore predominante. Il clima è influenzato a sua volta della cultura dell'organizzazione, vale a dire l'insieme coerente di assunti fondamentali che il gruppo di lavoro assume come il modo corretto di percepire, pensare e agire. Se si considera che la cultura dell'organizzazione descrive cos'è l'organizzazione stessa, il clima è un'indicazione di cosa i dipendenti sentono e credono sia l'organizzazione. Di conseguenza, il clima rappresenta le percezioni dei dipendenti riguardo l'organizzazione stessa e la produttività aziendale.

Quanto all'area valoriale va detto che l'esperienza lavorativa, al pari di altre esperienze umane, è un "oggetto" denso di significato, ricco di valenze individuali e collettive, alla cui definizione concorrono sia le concezioni valoriali sia la storia delle interazioni con soggetti e contesti specifici. Le

concezioni del lavoro possono essere considerate discendenti da un sistema interpretativo articolato. Al livello più alto di astrazione e generalizzabilità si collocano i valori, definibili come un insieme a elevata stabilità temporale di costrutti ipotetici che delineano cosa idealmente dovrebbe essere il lavoro rispetto agli scopi che è possibile-auspicabile perseguire attraverso esso (finalismo); cosa dovrebbe essere giusto fare (etica del lavoro) e cosa attiva, direziona e sostiene l'azione ovvero la inibisce definendo priorità e scelte (valenza motivazionale). I valori rappresentano pertanto dei parametri di valutazione delle azioni in relazione a standard sociali (efficacia dell'agire come "soggetto sociale competente") ed etici (giudizio circa la bontà, equità, correttezza, adeguatezza, ecc. dell'azione). I valori lavorativi possono essere definiti come parametri di valutazione ideale circa la desiderabilità di alcuni aspetti del lavoro, quali le sue caratteristiche stabili (es. retribuzione, condizioni, grado di autonomia ecc.) o i risultati che tramite esso è possibile raggiungere (es. soddisfazione, prestigio ecc.) (Schwartz 1992; Grimaldi 2007).

In una recente indagine (Inapp *et al.* 2024) i risultati ottenuti sono particolarmente interessanti e mostrano un cambiamento di prospettiva, in linea con quanto riportato dalla letteratura più recente. Le persone, infatti, intendono sempre più il lavoro come progetto di vita e non più solo come strumento di guadagno, mettendo al centro la qualità di vita. Le persone cercano molto più di un posto di lavoro e di guadagno, cercano situazioni in cui possono realizzare sé stessi dal punto di vista umano e professionale. Cercano ambienti di lavoro dove si sentano in armonia per ciò che riguarda i valori alla base del proprio vivere, dove vi sia il rispetto della persona umana, che si manifesta nell'equilibrio vita-lavoro, dove vi sia rispetto per ogni tipo di diversità (*gender equality*), dove si sentano 'ingaggiati' dal progetto professionale come parte attiva, come protagonisti e non solo forza lavoro. L'indagine evidenzia che l'importanza maggiore è attribuita alla possibilità di svolgere un'attività che piace, che offra l'opportunità di crescita a livello personale e professionale, che implementi le capacità e le competenze. Prevale quindi l'area della realizzazione e dell'accrescimento del sé; a seguire l'area della sicurezza e della stabilità assume un'importanza significativa. In conclusione, i risultati ci indicano che la cultura del lavoro si sta modificando passando, da una valenza che vedeva la sicurezza e la retribuzione come fattori predominanti, a una valenza che vede nel lavoro strette connessioni con il proprio progetto di vita: occasione di crescita e di valorizzazione del sé. Questi dati sono assolutamente in linea con quanto emerso da una recente ricerca dell'AIDP (Associazione italiana direzione del personale) che studiando il fenomeno delle dimissioni volontarie dal lavoro evidenzia che una motivazione importante era che il lavoro non soddisfaceva le aspettative di crescita professionale: così pur rimanendo fondamentali la sicurezza e la retribuzione, queste da sole non riescono più a giustificare l'idea che si possa continuare a lavorare in un ambiente considerato 'ostile'. Le persone e in particolare le nuove generazioni si aspettano di ricevere il giusto riconoscimento in termini di merito, ma guardano al lavoro anche come a un luogo dove stare bene e in cui vivere un clima relazionale sano ed equilibrato. Si rafforza così la convinzione che il lavoro non può essere solo un mezzo di sostentamento, ma un'attività attraverso la quale realizzarsi come persone. Indicazioni, queste, preziose per la programmazione di nuove politiche di valorizzazione delle risorse umane e quindi fondamentali per impostare un lavoro di gestione dei gruppi di lavoro per il leader. Un altro valore importante nelle organizzazioni oggi e che un buon leader deve perseguire è quello dell'inclusione. Per decenni e fino a poco tempo fa sono prevalsi concetti come inserimento e integrazione in relazione a soggetti che per diversa tipologia erano considerati fragili e/o diversi. Se il termine "inserimento" fa riferimento ad un processo sommativo per cui si aggiunge qualcuno in un dato insieme senza con questo preoccuparsi di quali scambi relazionali e comportamentali si attivano all'interno del gruppo, con il termine "integrazione" si assiste al passaggio concettuale per cui si pone enfasi alle interrelazioni tra i diversi soggetti del gruppo sociale che non viene più considerato la somma delle parti ma qualcosa di più complesso. Il termine "integrazione" si arricchisce così di variabili più soggettive e personali che intervengono nel processo. Con il termine "inclusione" l'accento si sposta su variabili più contestuali: è la società, il gruppo dominante che deve riorganizzarsi cogliendo le differenze e valorizzandole per integrare le diverse istanze personali al fine di promuovere benessere collettivo, collaborazione e cooperazione. L'inclusione diventa quindi una priorità per le organizzazioni pubbliche e private che siano e rappresenta un processo dinamico e democratico in cui le differenze diventano ricchezze. Si tratta di un processo multidimensionale che mira a creare le condizioni per una piena ed attiva partecipazione da parte di ogni membro alla vita lavorativa, anche a livello di processi propositivi e decisionali. Includere è quindi molto più complesso che integrare e presuppone impegno e solidarietà da parte di tutti, ma il ruolo del leader nel facilitare questo processo è decisivo (Grimaldi 2020; 2022).

Andando ora alle terza area di intervento individuata "Gestione e promozione del cambiamento" molteplici sono i segnali di interesse

e di approfondimento in questa direzione, a testimonianza di un dibattito a livello nazionale ed europeo vivace e aperto che vede come il mondo del lavoro e delle imprese segnala un crescente dinamismo di professionalità e di competenze a fronte della necessità di aggiornare o ridisegnare gli attuali sistemi di identificazione, classificazione e monitoraggio delle professioni. Anche il linguaggio utilizzato per descrivere il cambiamento (*professionals*, gestori di processo, ruoli trasversali, lavoratori della conoscenza, lavoratori invisibili, ecc.), se da un lato vuole evidenziare un livello di innovazione, dall'altro rappresenta bene l'indeterminatezza e l'instabilità attribuibili agli attuali sistemi. La complessità di un cambiamento culturale e di "sistema" nell'ambito delle politiche del lavoro si accompagna a contesti organizzativi sempre più avanzati e competitivi in grado di ricercare formule di analisi, valutazione e gestione delle risorse umane volte a rinforzare politiche di *flexicurity*. Gestire i cambiamenti oltre che promuoverli e anticiparli è un compito sempre più rilevante per un leader che deve adottare nuovi processi senza rallentare il team o l'organizzazione e cercando di superare le resistenze che di solito, ci ostacolano nell'abbandonare il "vecchio modo" di fare le cose, anche se quello nuovo è obiettivamente migliore. Un piano di gestione del cambiamento efficace è un processo complesso che rende consapevole il team del valore del nuovo cambiamento senza stravolgere i processi lavorativi abituali. Un processo efficace di gestione del cambiamento deve aiutare a dare impulso alla azienda, ridurre la resistenza, aumentare la probabilità di riuscita, ridurre l'eventuale impatto negativo del cambiamento tenendo conto delle competenze del team. Di interesse più recente, e sul quale si concentra molta letteratura tecnico-specialistica, è l'idea che nel corso della vita i lavoratori possono raggiungere opportunità concrete sulla base di un insieme mutevole di loro risorse e capacità. Il punto cruciale del concetto di capacità risiede nella combinazione di vari significati del verbo "potere" che si riferisce a tre aspetti: essere in grado di; avere l'opportunità di; essere facilitato e ammesso a. Per comprendere la situazione di un individuo e sviluppare interventi utili, è quindi importante valutare ciò che un individuo può fare o è in grado di fare, e non solo valutare quello che lui/lei fa in realtà. Ciò richiede, da un lato, un contesto socioeconomico che faciliti questo potenziale di risorse individuali e, dall'altro, un atteggiamento aperto alle opportunità che si presentano. L'attuale situazione socioeconomica e più in generale, la situazione culturale e quella sociale che stanno attraversando i Paesi dell'occidente, impongono di prefigurare misure e investimenti innovativi per favorire la crescita e la competitività e questo non può prescindere dalla gestione e dalla promozione di cambiamenti innovativi: cambiamenti nei processi, nei contenuti, nelle metodologie e negli strumenti del lavoro.

Infine per quanto concerne la quarta area di intervento individuata in relazione al lavoro del leader "Promozione e sviluppo delle reti nazionali e internazionali", ci riferiamo a tutte quelle attività che hanno a che fare con l'interazione e il raccordo con i soggetti economici/istituzionali che operano sul territorio nazionale e internazionale, raccordi che possono essere informali o formali attraverso la stipula di convenzioni e protocolli di intesa con i soggetti pubblici e privati che operano nel mercato del lavoro in quello specifico settore o in settori analoghi e complementari. Tale raccordo assume certamente anche la funzione di collegamento continuo ed efficace tra le imprese ai fini di una pronta risposta alla domanda di innovazione del mercato del lavoro. A tal riguardo importanti sono le attività di:

- ricerca e analisi degli indicatori socioeconomici;
- costruzione di piani di intervento (obiettivi, progetti, azioni, verifiche);
- analisi delle attività e dei risultati delle attività dei vari soggetti operanti nel mercato del lavoro;
- progettazione di prototipi, modelli di struttura e standard di servizi;
- progettazione di linee di servizi e di sperimentazione di attività innovative:
- verifica, valutazione e monitoraggio degli interventi (azioni/servizi) erogati;
- marketing e supporto ai fini del conseguimento degli obiettivi di cui sopra.

Prima di concludere questa breve dissertazione ci sembra un utile esercizio presentare, sulla base delle evidenze empiriche riscontrate e in considerazione delle stesse sollecitazioni dei testimoni privilegiati, un modello di competenze per il leader ai tempi dell'industria 5.0 che ci piacerebbe possa essere oggetto di un prossimo lavoro di verifica e che possa costituire la base per un modello formativo da sperimentare nei reali contesti applicativi. Siamo infatti convinti che se da un lato, per un lavoro sistemico che operi in una efficace logica di rete con tutti gli altri soggetti territoriali implicati nelle azioni di politica innovativa del lavoro, è necessario perseguire delle finalità strategiche (ottimizzazione delle funzioni, condivisione degli obiettivi, integrazione dei servizi dove siano definite modalità di coordinamento e interazione ma anche competenze specifiche e confini chiari), allo stesso tempo tali finalità devono costituire il terreno su cui fondare una nuova professionalità per i leader, se si vuole che effettivamente siano considerati facilitatori/gestori del cambiamento e una conseguente nuova

architettura dei processi formativi a loro dedicati. Se come sottolineato da più parti il ruolo del leader deve avere carattere di "dinamicità propositiva", in altri termini un leader 5.0 deve assumere il ruolo principale di "agente del cambiamento" sia a livello individuale sia a livello collettivo/sociale, allora è necessario impostare processi formativi dove possano essere sperimentate tecniche e metodologie per anticipare strategicamente i processi innovativi. L'utilizzazione del profilo del leader in contesti molto complessi e diversificati richiede l'ancoraggio a modelli, approcci e strumentazioni differenti e il possesso di requisiti e competenze personali e professionali eterogenee. Stante tali considerazioni e premesse, allo stato dei fatti, dalle interlocuzioni avviate e realizzate con la comunità di opinion leader intervistati ma anche tenendo conto di quanto evidenziato e raccomandato dalla letteratura sull'argomento è possibile individuare quattro macroaree di competenze che si collocano sempre più nell'intreccio tra le *hard* e le *soft skill* (figura 2) e dove le competenze che afferiscono all'intelligenza emotiva acquisiscono un peso determinante.

La prima macro-area fa riferimento alle **competenze relazionali-comunicative** comprende così la capacità di analizzare situazioni e problemi (personali, di gruppo e di contesto), di costruire una relazione interpersonale di aiuto e di sostegno, di garantire la valorizzazione e la crescita personale e professionale del singolo e dei gruppi anche per favorire clima positivo e benessere organizzativo, di gestire relazioni con contesti significativi di riferimento a seconda dell'ambito in cui opera, (aziende, imprese, scuole, servizi sociali, servizi del lavoro, università, ecc.), di leggere i bisogni e le dinamiche personali e di gruppo. Fondamentali sono la capacità di ascolto e di decodifica, l'attenzione alle fragilità, il riconoscimento della specificità dei target, la capacità di gestire il personale a livello individuale e di gruppo cercando allo stesso tempo di garantire la discrezionalità, la capacità di attivare il gruppo sia nelle attività caratterizzanti l'organizzazione nella quale si opera sia tra organizzazioni e istituzioni diverse dei sistemi che ruotano attorno all'ente di appartenenza. Tale macro-ambito di competenza sta assumendo una rilevanza crescente nelle realtà aziendali (o in collegamento con esse) per rispondere a nuove strategie di gestione delle risorse umane in funzione dei crescenti processi di ri-organizzazione dei contesti produttivi e/o di sostegno di percorsi individuali di sviluppo e carriera.
Esempi di competenze:

- tecniche di gestione del conflitto;
- competenze sui processi di comunicazione;
- capacità di ascolto attivo;
- capacità di assumere atteggiamenti e comportamenti flessibili;
- gestione del gruppo di lavoro e promozione di processi di coesione e interdipendenza
- promozione di scelte collettive e partecipative.

La seconda macroarea di competenza fa riferimento alle **competenze sui processi di analisi dei problemi, di progettualità e di innovazione di strategiche attività**. Questa area di competenze fa riferimento prevalentemente ai compiti specialistici di lettura e analisi della complessità culturale dei gruppi e delle organizzazioni in vista della definizione ed elaborazione dei futuri progetti strategici nonché della realizzazione e dell'operatività di tali percorsi futuri.
Esempi di competenze:

- capacità a leggere e trattare la complessità culturale dei singoli gruppi e delle organizzazioni;
- capacità ad affrontare l'indecisione e i problemi di resistenza;
- promozione di progetti sistemici e strategici;
- facilitare e gestire i processi di innovazione;
- capacità di *coping*, di problem solving e di resilienza
- capacità di valutazione della performance e di feedback**.**

La terza macroarea di competenze fa riferimento alle **competenze di lettura del contesto e di anticipazione dei cambiamenti** che comprende competenze di analisi del contesto, di definizione delle priorità, di programmazione degli interventi, di verifica dei risultati. Fa riferimento alla capacità di attivare i singoli sistemi e le reti territoriali, di gestire e coordinare azione integrate, di far dialogare istituzioni diverse, di progettare interventi per le risorse umane all'interno dei loro contesti lavorativi di appartenenza come risposta mediata fra bisogni dei lavoratori e richieste dell'organizzazione. Chiama in causa competenze giuridico-amministrative, di gestione del personale, di funzionamento delle organizzazioni. sottende capacità di analisi di sistemi complessi, di definizione delle priorità, di programmazione degli interventi, di verifica dei risultati, di interfaccia fra i livelli politico-programmatici di indirizzo e quelli tecnico-operativi di gestione delle attività e di consulenza sulle politiche.
Esempi di competenze:

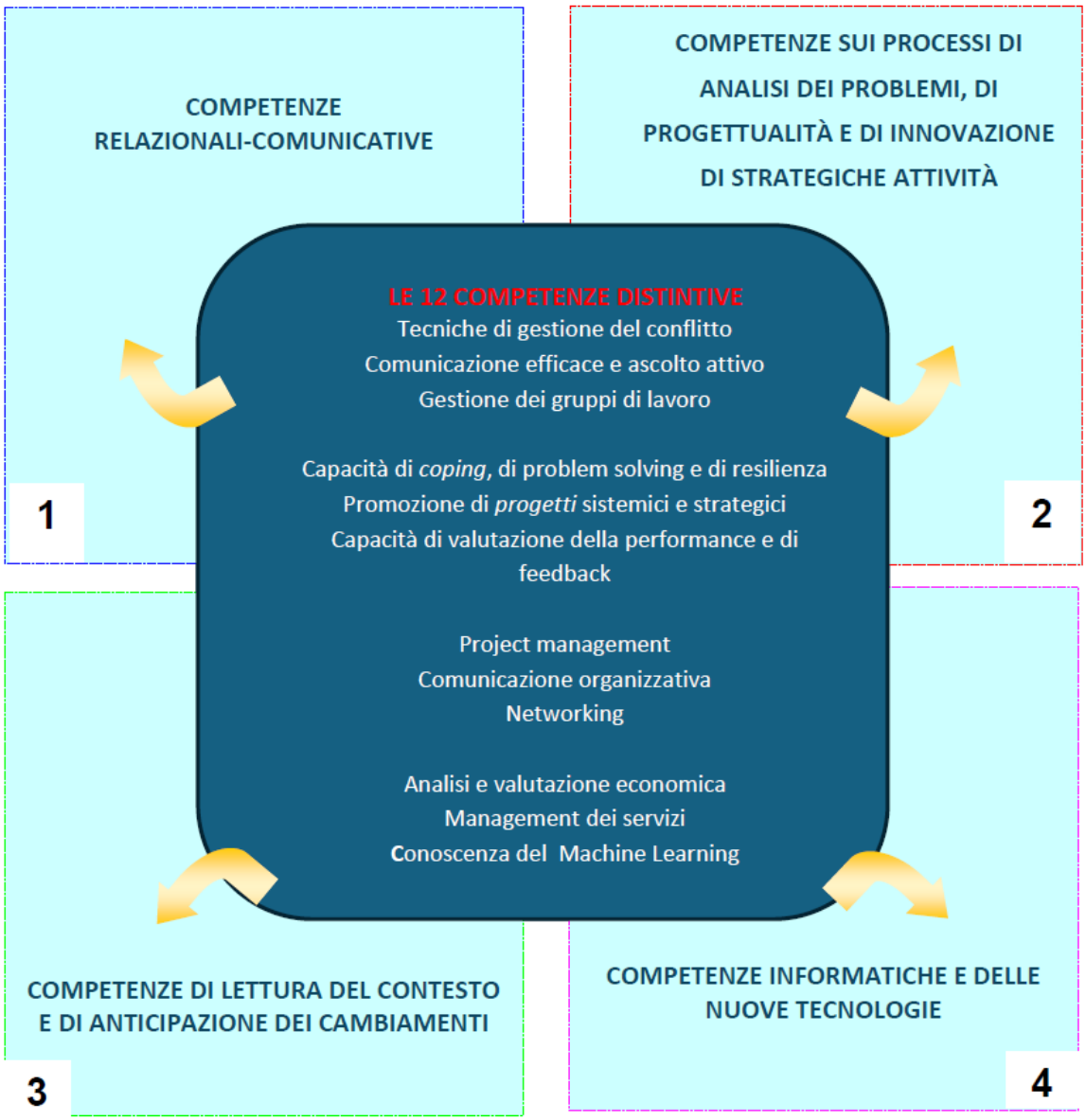
- project management (tecniche di gestione dei flussi informativi, di servizio e supporto progettuale attraverso l'integrazione di tecnologie, di analisi di procedure e strategie assunte anche dall'esterno);
- marketing e qualità (sviluppo e assicurazione qualità);
- gestione organizzativa sistemica (realizzazione e gestione del sistema di direzione per obiettivi);

- comunicazione organizzativa (tecniche di comunicazione e di micromarketing):
- negoziazione organizzativa (tecniche di gestione del conflitto a livello organizzativo);
- networking (gestione di sistemi e reti complesse).

La quarta macroarea fa riferimento alle **competenze informatiche e delle nuove tecnologie** con riferimento a competenze relative sia alle conoscenze da trasmettere sia alle metodologie comunicative di erogazione/diffusione di determinati contenuti sia alla consultazione di banche-dati e di altri canali utili per il reperimento delle informazioni da veicolare al destinatario finale e/o ad altri soggetti organizzativi. Sapersi muovere con destrezza con la tecnologia e l'informatica ormai è una necessità imprescindibile soprattutto alla luce dell'avvento dell'IA, anche per una buona gestione del personale, della finanza e del marketing così come delle problematiche e degli strumenti di gestione delle imprese e dei servizi a rete.
Esempi di competenze:

- capacità di analisi di statistiche descrittive e di validazione;
- capacità di analisi e valutazione economica (tecniche di analisi e lettura di dati economici e di investimento, di interpretazione di relazioni e piani gestionali);

Figura 2. Il profilo di competenze del leader



Fonte: elaborazione degli Autori

- conoscenza dei linguaggi *front-end* e *back-end*, nonché dei sistemi di gestione dei database;
- conoscenza del *Machine Learning;*
- management dei servizi.

Conclusioni

Il modello potrebbe trovare la sua sperimentale operatività in casi applicativi tratti da diversi settori, in cui la leadership 5.0 dimostra la capacità di trasformare l'Intelligenza artificiale in alleato del giudizio, non nel suo sostituto. Non si vogliono qui eludere i limiti dell'approccio, ma si assumono come punto di partenza per ulteriori sviluppi. La natura qualitativa e la dimensione contenuta del campione non consentono inferenze statistiche, ma offrono una base solida per un'agenda di ricerca futura, con verifiche quantitative, metriche longitudinali e confronti intersettoriali. Al tempo stesso, l'attenzione all'operatività mostra che il modello proposto è utilizzabile come riferimento concreto per la progettazione di formazione, governance e valutazione, purché trattato come mappa interpretativa e non come check list prescrittiva.

Le conclusioni ricompongono infine i piani teorico, empirico e propositivo, rispondendo alla domanda di ricerca iniziale. L'Intelligenza artificiale ridefinisce ruoli, aree d'azione e competenze della leadership non perché sostituisca l'umano, ma perché ristruttura il contesto cognitivo e organizzativo della decisione: moltiplica la disponibilità di dati, comprime i tempi, introduce nuove opacità e apre spazi inediti di possibilità e responsabilità. La leadership efficace riconosce potenza e fallibilità degli strumenti, protegge il giudizio umano, disegna confini e soglie, rende trasparenti ipotesi e metriche, promuove il confronto e legittima il dubbio come antidoto alla delega automatica. Le condizioni culturali e organizzative che consentono un'adozione generativa dell'IA sono il cuore della questione: coerenza tra valori dichiarati e pratiche, qualità delle conversazioni decisionali, cura del clima e del benessere, regole condivise e *accountability* diffusa. In questa logica la tecnologia torna a essere mezzo al servizio di fini che mettono al centro la dignità e lo sviluppo umano. La leadership 5.0, come emerge dal lavoro, è l'arte di umanizzare ciò che la tecnologia rende possibile, custodendo il nesso tra efficienza e senso, tra performance e giustizia organizzativa, tra innovazione e responsabilità. Se questo nesso è mantenuto vivo e praticato, l'Industria 5.0 diventa promessa mantenuta di lavoro migliore e decisioni più consapevoli; se viene trascurato, l'IA rischia di trasformarsi nell'ennesima occasione mancata. La differenza la fa, ancora una volta, la qualità della leadership: capacità di ascoltare, spiegare, collegare, proteggere, rendere conto e trasformare l'Intelligenza artificiale in alleato del giudizio, non nel suo sostituto.

Bibliografia

Alshaibani E., Bakir A., Al-Atwi A. (2025), The impact of leadership behaviors on organizational innovative performance and learning in AI-driven Industry 5.0 environments, *Development and Learning in Organizations: An International Journal*, 39, n.3, pp.18-21

Ancora A., Grimaldi A., Scarcella G. (2023), *Professione intervistatore: l'importanza della conduzione dell'intervista per la qualità del dato*, Inapp Paper n.46, Roma, Inapp

Avolio B. J., Bass B. M. (2004), *Multifactor leadership questionnaire manual*, Redwood City, Mindgarden

Burgess M. (2024), *The executive guide to artificial intelligence*, Palgrave Macmillan Cham

Chamakiotis P., Panteli N., Davison R. M. (2021), Reimagining e-leadership for reconfigured virtual teams due to COVID-19, *International Journal of Information Management*, n.60

D'Emilione M., Giuliano G., Grimaldi A. (2020), La collaborazione tra professionisti e operatori sociali nelle politiche a contrasto della povertà: il ruolo dell'equipe multidisciplinare, *Counseling Giornale Italiano di Ricerca e Applicazioni*, 13, n.2, pp.16-41

De Cremer D. (2022), With AI entering organizations, responsible leadership may slip!, *AI and Ethics*, n.2, pp.49-51

Di Fabio A., Bonfiglio A., Palazzeschi L., Gori A., Svicher A. (2023), Human capital sustain ability leadership: From personality traits to positive relational management, *Frontiers in Psychology*, n.14, Article 1110974

Di Fabio A., Gori A. (2021), Workplace relational civility and human capital sustainability leadership for sustainable development in organizations: Empirical evidence, *Counseling: Rivista Internazionale di Ricerca e Applicazioni*, 14, n.2, pp.32-40

Fiedler F.E. (1964), A Contingency Model of Leadership Effectiveness, *Advances in Experimental Social Psychology*, n.1, pp.149-190

Goleman D. (1998), What makes a leader?, *Harvard Business Review*, 76, n.6, pp.93-102

Grimaldi A. (2025), Soft skill: tra concezioni culturali e pratiche operative, in Remo Job e Marisa Michelini (a cura di), Competenze trasversali, dalla formazione al lavoro, RicercAzione Risorse n. 2, Trento, Iprase, pp.77-88

Grimaldi A. (2022), Cittadinanza attiva e inclusione sociale: intrecci di due frontiere complementari, in Michelini M., Tirelli G. (a cura di), *Educazione alla cittadinanza e costituzione*, Bologna, Brixia University Press, pp.65-78

Grimaldi A. (2020), Le competenze per l'occupabilità come risorsa strategica per il proprio progetto di vita in Pellerey M., Margottini M., Ottone E. (a cura di), *Dirigere sé stessi nello studio e nel lavoro. Competenzestrategiche.it: strumenti e applicazioni*, Roma, TrE-Press Editore, pp.193-204

Grimaldi A. (2010), Il documento Isfol sul bilancio di competenze/2, *Professionalità*, n.109, pp.44-55

Grimaldi A. (2009), Sulle competenze: Insidie e opportunità, *Counseling. Giornale Italiano di Ricerca e Applicazioni*, n.2, pp.255-270

Grimaldi A. (a cura di) (2007), *Bisogni, valori e autoefficacia nella scelta e nella ricerca del lavoro*, Roma, Isfol

Hersey P., Blanchard K. (1988), *Management of Organizational Behaviour,* Hoboken (US), Prentice Hall

House R.J., Mitchell T.R. (1977), Path-goal Theory of Leadership, *Journal of Contemporary Business*, 6, n. 3, pp. 81-97

Inapp, Ancora A., Grimaldi A. (a cura di) (2024), *L'orientamento tra nuovi bisogni e nuove competenze: un'indagine sulla popolazione giovanile*, Inapp Report n.50, Roma, Inapp

Jago A.G. (1982), Leadership: Perspectives in Theory and Research, *Management Science*, 28, n.3, pp.315-336

Laybats C., Tredinnick L. (2024), Ethics and artificial intelligence. The rise of the deep fake, *Business Information Review*, 41, n.2, pp.31-46

Liu C., Wolter C., Xian W., Jeschke J.M. (2020), Most invasive species largely conserve their climatic niche, *Proceedings of the National Academy of Sciences*, 117, n.38, pp.23643-23651

López-Figueroa J.C., Ochoa-Jiménez S., Palafox-Soto M.O., Hernandez Munoz D.S. (2025), Digital leadership: A systematic literature review, *Administrative Sciences*, 15, n.4, Article 129

Machado C., Davim J. P. (2023), *Industry 5.0: Creative and innovative organizations*, Cham, Springer

Oh C., Ki C. (2024), Can we build a relationship through AI?, *Public Relations Review*, 50, n.4, Article 102469

Rezazadeh A., Kohns M., Bohnsack R., António N., Rita P. (2025), Generative AI for growth hacking: How startups use generative AI in their growth strategies, *Journal of Business Research*, 192, Article 115320

Schein E. H. (2010), *Le forme dell'aiuto*, Milano, Raffaello Cortina Editore

Schein E. H. (1985), *Organizational culture and leadership*, San Francisco (CA), Jossey-Bass Publisher

Schwartz S. (1992), Universals in the content and structure of values: Theoretical advances and empirical tests in 20 countries, in Zanna M. (a cura di), *Advances in experimental social psychology*, New York, Academic Press, pp.1-65

Stogdill R. M. (1974), *Handbook of leadership: A survey of theory and research*, Free Press

Wagner J.A., Hollenbeck J.R. (1992), *Management of Organizational Behavior*, Englewood Cliffs (NJ), Prentice-Hall

Yukl G. (2013), *Leadership in organizations* (8th ed.), Boston, Pearson

Anna Grimaldi
grimanna56@gmail.com
Dottoressa di ricerca in psicologia sperimentale, dirigente di ricerca, già responsabile della struttura Inclusione sociale e Presidente del CUG dell'Inapp. Direttrice del corso di Sviluppo delle competenze per l'occupabilità (soft skill) presso l'università UnitelmaSapienza di Roma. Tra le pubblicazioni più recenti si segnalano: Soft skill: tra concezioni culturali e pratiche operative, *RicercAzione RISORSE*, n.2/2025; L'orientamento nella transizione scuola-università-lavoro: dallo stato dell'arte alle prospettive future, in Caputo A. (2023), *Pensare il futuro*, Franco Angeli.

Alfonso Monsurrò
alfonso.monsurro@unicredit.eu
Dottore in Strategic Management presso la LUISS Università di Roma. I suoi interessi si concentrano sui temi della leadership, della cultura organizzativa e delle trasformazioni generate dall'Intelligenza artificiale nei contesti produttivi e manageriali. Nel suo percorso di studi ha approfondito l'evoluzione dei modelli di leadership nell'ambito dell'Industria 5.0, con particolare attenzione al rapporto tra capitale umano, processi decisionali e tecnologie intelligenti.

L'istituzionalizzazione della partecipazione pubblica in Italia

Una proposta di periodizzazione dagli anni Ottanta

Vittoria Laino
Sapienza Università di Roma
e Universidade de Coimbra
(Dottoranda)

Il contributo esamina il processo di istituzionalizzazione della partecipazione pubblica in Italia dagli anni Ottanta. Una griglia teorica che pone in relazione i paradigmi della democrazia partecipativa, del governo aperto e dell'amministrazione condivisa orienta la ricognizione di 93 dispositivi di *hard* e *soft law*, sulla cui base è proposta una periodizzazione in cinque fasi. L'analisi evidenzia il passaggio da pratiche sperimentali e settoriali alla configurazione della partecipazione come campo di policy autonomo, con la progressiva emersione di un'infrastruttura multilivello.

The institutionalisation of public participation in Italy. A proposed periodisation from the 1980s onwards

The article examines the institutionalisation of public participation in Italy since the 1980s. A theoretical framework linking the paradigms of participatory democracy, open government and shared administration guides the mapping of 93 hard-law and soft-law instruments, on the basis of which a five-phase periodisation is proposed. The analysis highlights the transition from experimental and sector-specific practices to the establishment of public participation as a policy field in its own right, alongside the gradual emergence of a multilevel infrastructure.

DOI: 10.53223/Sinappsi_2026-02-13

Citazione
Laino V. (2026), L'istituzionalizzazione della partecipazione pubblica in Italia. Una proposta di periodizzazione dagli anni Ottanta, *Sinappsi*, XVI, n.2, pp.189-206

Parole chiave
Amministrazione condivisa
Democrazia
Sistemi di governance

Keywords
Shared administration
Democracy
Governance frameworks

Introduzione

La partecipazione pubblica rappresenta – come noto – un ambito plurale e semanticamente eterogeneo. In questo contributo, l'espressione designa il campo empirico delle pratiche e dei dispositivi attraverso cui la cittadinanza viene coinvolta in modo strutturato nei processi di azione pubblica, secondo "un'accezione più ristretta di partecipazione, intesa come relazionamento della società con le istituzioni che comporta un intervento di espressioni dirette della prima nei processi di azione delle seconde" (Bobbio e Pomatto 2007, 2). La democrazia partecipativa costituisce il principale paradigma teorico di riferimento attraverso cui tali forme di coinvolgimento vengono interpretate. L'attenzione è rivolta pertanto a quella famiglia di istituti riconducibili a spazi di partecipazione attivati e regolati dalle istituzioni – i cosiddetti *invited spaces*, nei quali "le persone […] sono invitate a partecipare da autorità di vario tipo" (Gaventa 2006, 26)[1] e ove "la partecipazione non avviene spontaneamente, ma viene avviata" (Wilcox 2003, 4).

Il contributo si propone di ricostruire le principali traiettorie e discontinuità del processo di istituzionalizzazione della partecipazione pubblica in Italia, attraverso una ricognizione multilivello dei dispositivi di *hard* e *soft law* che hanno progressivamente regolato e strutturato le

1 Salvo diversa indicazione, le traduzioni dei passi citati sono dell'Autrice.

modalità di coinvolgimento della cittadinanza nei processi di elaborazione delle politiche. Assunti come tracce normative dei mutamenti nelle forme di riconoscimento e integrazione degli apporti civici nei processi di policy-making, tali dispositivi offrono una lettura necessariamente parziale del processo di istituzionalizzazione. La ricostruzione intende dunque contribuire alla comprensione della trasformazione della partecipazione pubblica da pratica sperimentale e settoriale a dimensione riconoscibile e progressivamente regolata dell'azione pubblica.

La periodizzazione proposta prende avvio dagli anni Ottanta, pur riconoscendo l'esistenza di precedenti tentativi di istituzionalizzazione – in particolare durante gli anni Settanta – qui non ricostruiti sistematicamente per ragioni di delimitazione del campo di indagine.

1. Coordinate concettuali

La partecipazione pubblica si articola in pratiche caratterizzate da diversi gradi di formalizzazione e intensità, operanti su molteplici livelli amministrativi e lungo l'intero processo di policymaking.

Sebbene – nell'ambito degli istituti annoverabili nel dominio della democrazia partecipativa – la responsabilità della decisione finale su un determinato intervento rimanga di regola in capo alle istituzioni pubbliche competenti[2] e gli esiti partecipativi siano spesso non vincolanti sul piano giuridico, ciò non predetermina l'intensità del rapporto. Questa distinzione tra titolarità formale della decisione e intensità sostanziale del coinvolgimento contribuisce a rendere strutturalmente tesa la relazione tra innovazioni partecipative e architettura rappresentativa, che si manifesta nella variabilità degli impegni assunti dai decisori e nelle modalità con cui viene definita la capacità di influenza degli esiti partecipativi.

Ai fini della delimitazione dell'oggetto di analisi, le pratiche in oggetto sono qui intese come forme di interazione tra cittadinanza e istituzioni pubbliche, strutturate attraverso procedure previamente organizzate e relativamente stabili, che possono includere a diverso titolo soggetti esperti; le parti, nel reciproco riconoscimento in qualità di legittimi attori, interagiscono con intensità variabile lungo un continuum che si estende dalle forme consultive a quelle collaborative.

La partecipazione pubblica è assunta come campo empirico degli strumenti e delle pratiche attraverso cui la cittadinanza interviene in modo strutturato nei processi di azione pubblica e il dominio della democrazia partecipativa acquista forma operativa. Entro tale paradigma generale sono collocati due ulteriori sotto-paradigmi – il governo aperto nella sua componente partecipativa e collaborativa e l'amministrazione condivisa – quali declinazioni specifiche corrispondenti a modalità e gradi diversi di coinvolgimento civico. In funzione descrittiva rispetto alla documentazione illustrata nelle sezioni successive, si considerano pertanto tre cornici teoriche tra loro correlate, che offrono una griglia minima funzionale a inquadrare i dispositivi considerati nel processo italiano di istituzionalizzazione.

Il paradigma della democrazia partecipativa rappresenta un dominio democratico fondamentale entro il quale possono essere ricompresi esercizi di "inclusione della cittadinanza" (Moro 2022, 147) nei processi che interessano l'elaborazione di corsi d'azione nello spazio pubblico, dalla definizione e attuazione di interventi a forme collaborative di svolgimento di attività di interesse generale[3].

Sul piano empirico le esperienze risultano diversificate per forme e procedure, delineando il carattere ibrido di un campo "di sperimentazione di forme e procedure di riforma della democrazia volte a includere più strettamente ed efficacemente i cittadini nel processo politico, specialmente con riferimento alla formazione delle policy che investono direttamente le loro vite" (Moro 2009, 2), che comprende pratiche eterogenee accomunate da una tensione inclusiva dalle ricadute pragmatico-operative[4]. In rapporto alla democrazia deliberativa, la democrazia partecipativa può essere assunta come ambito in cui si registra un avanzamento sostanziale del ruolo della cittadinanza nei processi decisionali pubblici; la deliberazione, intesa come "procedura orizzontale" (Bobbio *et al*. 2017, 128), costituisce una componente qualificante degli

2 Con riferimento agli istituti orientati alla formazione delle politiche, si osserva che «la democrazia partecipativa si muove in un ambito che, sul piano giuridico, potremmo definire "consultivo". La decisione finale spetta sempre alle istituzioni della democrazia rappresentativa […] Se poniamo la questione in termini di "capacità di influenza" piuttosto che di "vincolo giuridico", la questione chiave è quella di analizzare quale tipo di impegno si assumono i promotori politici» (Bobbio e Pomatto 2007, 26).

3 Una tale definizione consente di ampliare l'orizzonte teorico e operativo della democrazia partecipativa, evitando di limitarla agli istituti che coinvolgono la cittadinanza solo nei processi antecedenti alla decisione politica.

4 Le forme strutturate di democrazia partecipativa possono includere, ad esempio, bilanci partecipativi, assemblee di cittadini, *consensus conferences*, *deliberative pollings*, dibattiti pubblici, giurie dei cittadini, *town meetings*, ciascuna caratterizzata da una struttura prestabilita. È possibile poi individuare un'ampia gamma di tecniche partecipative combinabili in molteplici modi a seconda del contesto e degli obiettivi. Cfr. Laino 2025, 35-43.

esercizi e, ove adeguatamente realizzata, può contribuire alla qualità dell'interazione, pur non trovando necessariamente piena attuazione in tutte le forme di coinvolgimento riconducibili al paradigma[5].

Il paradigma del governo aperto, affermatosi al livello internazionale all'interno dei dibattiti sulla trasparenza amministrativa e successivamente ampliatosi sino a comprendere la partecipazione e la collaborazione tra amministrazioni e cittadinanza, trova un riferimento istituzionale nell'iniziativa multilaterale Open Government Partnership, lanciata nel 2011.

Nella sua configurazione generale, il governo aperto può essere inteso come un paradigma di governance multidimensionale, definito come "una cultura di governo che promuove i principi di trasparenza, integrità, responsabilità e partecipazione degli stakeholder a sostegno della democrazia e della crescita inclusiva" (OECD 2023, 3). Ai fini del presente contributo, l'attenzione è circoscritta alla sua componente partecipativa e collaborativa, che pone in risalto la permeabilità delle istituzioni e la necessità di integrare le competenze dei cittadini per migliorare la qualità dell'elaborazione delle politiche. E dunque l'*openness of government* "non riguarda soltanto l'apertura in termini informativi […] ma anche l'apertura in termini interattivi" (Meijer *et al*. 2012, 10), e può essere intesa come "il grado in cui i cittadini possono monitorare e influenzare i processi governativi attraverso l'accesso alle informazioni pubbliche e agli spazi decisionali" (ivi, 13), evidenziando la duplice natura – informativa e partecipativa – del paradigma[6].

La componente interattiva si inscrive nel processo di ripensamento delle modalità di governo che ha favorito l'emergere, accanto al modello tradizionale di tipo gerarchico-burocratico, di forme decisionali cooperative e reticolari miste pubblico-private, riconducibili alla *network governance* (d'Albergo 2014, 178; Moini 2012, 66-67). Tale evoluzione ha aperto la strada a forme di governance collaborativa, nelle quali "una o più amministrazioni pubbliche coinvolgono direttamente attori non statali in un processo decisionale collettivo, formale, orientato al consenso e di natura deliberativa, finalizzato alla definizione o all'attuazione di politiche pubbliche o alla gestione di programmi e beni pubblici" (Ansell e Gash 2008, 544).

Fondato sul principio di sussidiarietà costituzionale (art. 118, u.c. Cost.) (Rescigno 2002), il paradigma dell'amministrazione condivisa è generalmente definito come un modello organizzativo che consente a cittadini e amministrazioni pubbliche di svolgere, su un piano paritario, attività di interesse generale concernenti la cura, la rigenerazione e la gestione condivisa dei beni comuni[7]. Nella griglia concettuale adottata, esso costituisce la declinazione maggiormente operativa della democrazia partecipativa.

Nel contesto italiano, il modello si innesta su una specifica traiettoria costituzionale, in cui alle istituzioni pubbliche è attribuito il compito di "favorire l'autonoma iniziativa dei cittadini" nello svolgimento di attività di interesse generale. La sussidiarietà costituzionale ha rappresentato in questa prospettiva una leva per il processo di istituzionalizzazione dell'amministrazione condivisa quale forma collaborativa della partecipazione pubblica, la cui regolamentazione ha "lo scopo di dare attuazione al principio costituzionale […] facendolo penetrare in profondità nell'azione e nell'organizzazione amministrativa comunale"[8]. Ne deriva un'impostazione pragmatico-pluralista orientata a riconfigurare la funzione amministrativa come ambito relazionale, fondato sull'interazione e sulla corresponsabilità tra attori istituzionali e cittadini attivi, nella prospettiva secondo cui "la promozione dell'interesse generale e la partecipazione alla funzione pubblica fossero un diritto-dovere di tutti" (Pezzana 2023, 29).

La *ratio* culturale e giuridica del modello,

5 Con riferimento alla configurazione procedurale di tali esercizi, Bifulco osserva che le forme della democrazia partecipativa danno espressione alle teorie della democrazia deliberativa nella misura in cui soddisfano «i due requisiti della effettiva partecipazione di tutti gli interessati e della natura deliberativa della partecipazione […] Mi pare che, *per il modo stesso in cui sono organizzate, le forme concrete della dp si confermino essere applicazioni delle teorie della dd*» (Bifulco 2009, 5). L'orientamento deliberativo delle tecniche non assicura, tuttavia, l'effettiva qualità deliberativa delle pratiche partecipative.

6 Si vedano Open Government Partnership (2025) e OECD (2016).

7 Per un'analisi complessiva del modello di amministrazione condivisa, si vedano i diversi contributi raccolti nel volume curato da Arena e Bombardelli (2022). Si veda inoltre: Caianiello L., Amministrazione condivisa, in Glossario dell'amministrazione condivisa, *Labsus*, https://www.labsus.org/glossario-dellamministrazione-condivisa (consultato il 05/10/2025).

8 Muzi L., Comune di Bologna, *Regolamento sulla collaborazione tra cittadini e amministrazione per la cura e la rigenerazione dei beni comuni urbani*, deliberazione del Consiglio comunale 19 maggio 2014, n. 172, disponibile su Labsus: https://www.labsus.org/2014/11/bologna-delibera-consiglio-comunale-19-maggio-2014-n-172-regolamento-sulla-collaborazione-tra-cittadini-e-amministrazione-per-la-cura-e-la-rigenerazione-dei-beni-comuni-urbani/ (consultato il 05/10/2025).

"applicabile a ogni ambito di intervento pubblico" (Arena 2022, 9), punta dunque a ricomporre la distanza tra amministrazione pubblica di riferimento e cittadinanza, delineando "un modo nuovo di amministrare" (Arena 2004, 1) in discontinuità rispetto alla *standard view* del modello amministrativo della gestione pubblica centralizzata. L'amministrazione condivisa, tuttavia, non sostituisce tale modello, ma gli 'si affianca'[9] introducendo una corresponsabilità regolata, nella quale l'ente pubblico attiva le condizioni istituzionali per la collaborazione, mentre i cittadini contribuiscono all'azione di interesse generale.

Si ritiene dunque che partecipazione e collaborazione, pur non essendo sinonimi, non rinviino a famiglie concettuali e operative eterogenee, bensì a gradi diversi di coinvolgimento all'interno della stessa traiettoria di inclusione civica, che va dai meccanismi di consultazione pre-decisionali fino a forme di co-gestione e co-valutazione.

La governance partecipativa rinvia a configurazioni istituzionalizzate attraverso le quali attori non statali sono coinvolti in processi collettivi che possono riguardare non solo la definizione degli indirizzi, ma anche la gestione di programmi e beni comuni; in tali assetti, la partecipazione può assumere forme collaborative e intrecciarsi con la co-produzione di valore pubblico, dando luogo alla condivisione di responsabilità. In questo quadro, all'interno del paradigma generale della democrazia partecipativa, il governo aperto e l'amministrazione condivisa ne rappresentano due declinazioni: il primo, nella sua componente partecipativa e collaborativa, è incentrato sull'apertura interattiva delle istituzioni e sull'inclusione degli stakeholder; la seconda è orientata alla produzione e alla gestione congiunta di beni e servizi, sulla base del principio di sussidiarietà costituzionale.

Gli atti censiti e sistematizzati in questo lavoro – sulla cui base è costruita la periodizzazione – sono qui assunti come riconducibili, nelle rispettive declinazioni formali e operative, alle cornici teoriche summenzionate, che pertanto vengono trattate congiuntamente ai fini dell'analisi.

2. Dimensioni dell'istituzionalizzazione

Con il termine *istituzionalizzazione* viene generalmente inteso l'"incorporare nuove forme di partecipazione nelle strutture e nei processi decisionali politici legalmente fondati di una comunità o di una società […] [iter che] comporta la persistenza e la riproduzione di un modello di comportamento o di una procedura che procede indipendentemente dalle caratteristiche, dagli interessi e dalle preferenze degli individui che ne sono coinvolti" (Brunazzo 2017, 838-839). Si tratta dunque di un processo volto a regolare la partecipazione attraverso "norme e procedure, [al fine di] farla diventare una routine abituale […]. Il grado di istituzionalizzazione delle pratiche partecipative varia in ragione della loro sistematicità (piuttosto che episodicità), legittimità intrinseca e spersonalizzazione (ovvero il non essere connesse a specifici individui)" (Lewanski 2016, 133).

Per ciò che concerne il processo di regolamentazione istituzionale della partecipazione, è possibile distinguere dispositivi di *hard law* e di *soft law*[10], i quali hanno contribuito – su più livelli amministrativi – alla costituzione e alla sistematizzazione formale e prescrittiva degli istituti partecipativi.

Formalizzazione e istituzionalizzazione

La distinzione tra formalizzazione e istituzionalizzazione rappresenta un passaggio analitico preliminare utile alla comprensione di come pratiche partecipative, inizialmente episodiche o sperimentali, vengano progressivamente incorporate entro dispositivi riconoscibili e riproducibili dell'azione pubblica.

La formalizzazione – che può anche essere di natura extra-istituzionale, ossia operata da soggetti esterni non appartenenti agli apparati amministrativi[11], secondo finalità differenti – consiste in un'operazione di standardizzazione e sistematizzazione che rinvia al "grado in cui le informazioni vengono incanalate in un certo modo, lasciando maggiore o minore spazio alla comunicazione aperta" nell'ambito di una pratica partecipativa (Newig *et al.* 2008, 424). Ne deriva

9 Caianiello L., Amministrazione condivisa, in *Glossario dell'amministrazione condivisa*, *Labsus*, https://www.labsus.org/glossario-dellamministrazione-condivisa (consultato il 05/10/2025).

10 La distinzione tra *hard law* e *soft law* si fonda principalmente sul diverso grado di vincolatività degli strumenti considerati. Gli atti di *soft law*, pur non essendo dotati di piena forza obbligatoria, sono comunque in grado di orientare le prassi applicative e di produrre effetti giuridici indiretti, contribuendo ad attenuare la tradizionale separazione tra le due categorie (Angeli 2025). Nel presente contributo, la distinzione è adottata in senso funzionale: sono ricondotti alla *hard law* anche gli atti amministrativi di pianificazione e programmazione formalmente approvati che, pur non costituendo fonti normative in senso stretto, producono effetti conformativi sull'azione amministrativa o condizionano l'attuazione degli interventi e l'accesso alle risorse.

11 Come accademici, organizzazioni della società civile ed esperti di policy.

che un metodo maggiormente formalizzato "funge da filtro, permettendo solo a informazioni di una particolare natura di passare, escludendone altre" (ivi, 425). Il processo risulta funzionale a una cristallizzazione sistematica degli oggetti capace, nella migliore delle ipotesi, di contribuire alla formazione di una cultura della partecipazione pubblica e, in una lettura più prudente, di accrescere la consapevolezza e la conoscenza della materia.

L'istituzionalizzazione, d'altra parte, riguarda il processo di incorporazione delle pratiche entro le strutture e le procedure decisionali politicamente legittime, orientato a far sì che la partecipazione diventi "un'attività riconosciuta, routinizzata e sostenibile" (Randma-Liiv 2023, 330). Una dimensione significativa di tale processo è quella legata al *policy design*; l'istituzionalizzazione della partecipazione non coincide difatti con la sola proliferazione di riferimenti normativi, ma implica la progettazione di strumenti dedicati, che definiscano in modo esplicito chi può partecipare, quando, come e con quali impegni di ritorno da parte dei decisori pubblici.

In questo senso la formalizzazione si configura come un momento logicamente antecedente – sebbene non necessariamente cronologico – mediante il quale si opera una cristallizzazione concettuale e normativa di procedure e istituti 'fluidi' e frammentari. La formalizzazione – si sostiene – esercita una funzione abilitante rispetto all'istituzionalizzazione, creando le condizioni semantiche entro cui diviene possibile, sebbene non inevitabile, il riconoscimento istituzionale di pratiche precedentemente non codificate o scarsamente sistematizzate.

Tuttavia, non deve ritenersi che ogni processo di formalizzazione conduca automaticamente a istituzionalizzazione, né che ogni istituzionalizzazione sia preceduta da una piena formalizzazione. E formalizzazione e istituzionalizzazione non conducono – naturalmente – alla produzione di output in modo automatico e non costituiscono di per sé garanzia di trasformazione dei processi di policy-making. Tale disaccoppiamento rimanda alla distinzione tra dimensioni formali e informali dell'istituzionalizzazione: l'incorporazione entro dispositivi, infatti, resta condizionata anche dalla compresenza di una dimensione informale fatta di valori, credenze e orientamenti di decisori e apparati (Randma-Liiv 2023, 331).

Il potenziale disallineamento tra adozione e incidenza sul policy-making dipende anche dal fatto che la codificazione delle regole non elimina né l'ambiguità interpretativa né la discrezionalità nell'applicazione (Mahoney e Thelen 2010, 10-11). Quando ciò avviene, l'istituzionalizzazione rischia di ridursi a un meccanismo esposto a pratiche di svuotamento semantico funzionali alla riproduzione di assetti consolidati, nonché a un esito di burocratizzazione per cui il processo, lungi dal rafforzare la capacità di incidenza della cittadinanza nei circuiti decisionali, tende a inscrivere la partecipazione entro logiche amministrative e procedurali che ne attenuano la portata trasformativa, esponendola al rischio di ritualizzazione simbolica (Lewanski 2016, 171)[12].

La transizione da pratica episodica a sistema strutturato e riconosciuto rimane un processo non lineare; e più in generale la governance può essere intesa come un "processo complesso e continuo di interpretazione, conflitto e attività, che produce configurazioni del governo sempre mutevoli" (Bevir 2010, 437). Le iniziative partecipative possono dunque restare esperimenti formalmente ordinati ma politicamente poco trasformativi e con poca significatività per la democrazia (Michels e Binnema 2019, 766), confermando il carattere conteso dell'istituzionalizzazione.

Traiettorie multilivello

Pur riconoscendo il livello locale come campo privilegiato di esercizio dei sistemi collaborativi (Fung e Wright 2003; Yi e Qiu 2025), negli ultimi vent'anni il lessico della governance partecipativa ha assunto un ruolo sempre più centrale nelle agende politiche sovranazionali, con conseguenti sperimentazioni[13]. Nel contesto europeo, tale dinamica è interrelata a quella della crisi di legittimazione che attraversa le democrazie contemporanee, frequentemente descritta attraverso indicatori di disaffezione, calo della fiducia e indebolimento dei canali rappresentativi tradizionali (Pudar Drasko e Fiket 2026, 2).

L'apertura collaborativa si è dispiegata attraverso un intreccio di atti e disposizioni di *hard law* (ad es. Carta europea dell'autonomia locale; art. 11 del Trattato sull'Unione europea; Convenzione di Aarhus) e di *soft law* (ad es. *White Paper on European Governance*; *Code of Good Practice for Civil Participation in the Decision-Making Process*;

12 Anche al livello europeo è stato espresso tale rischio: "Mentre nei documenti e nei discorsi ufficiali dell'UE si fa spesso riferimento a una 'svolta partecipativa' […], la sua realizzazione pratica sembra essere più difficile da osservare [...] tali strumenti di politica partecipativa possono essere considerati come un elemento di una 'riforma simbolica'" (Boussaguet 2016, 119-120).

13 Come la nota Conferenza sul futuro dell'Europa (2021-2022).

Guidelines for civil participation in political decision making; Raccomandazione (UE) 2023/2836), riconducibili a differenti istituzioni sovranazionali e intergovernative[14].

La *soft law* anche in questo caso assume un ruolo propulsivo nel delineare un processo di istituzionalizzazione 'morbida', alimentato da atti (ad es. raccomandazioni, linee guida, comunicazioni) che hanno agito orientando il comportamento delle istituzioni e diffondendo pratiche di dialogo e collaborazione tra cittadini e amministrazioni pubbliche. Essa può essere interpretata come una categoria di dispositivi che opera simultaneamente su due piani: quello della regolamentazione, in quanto produce un quadro di riferimento che orienta i comportamenti amministrativi pur senza vincoli giuridici diretti, e quello culturale, poiché contribuisce a normalizzare la partecipazione come componente intrinseca della "buona governance", favorendone la percezione come parte legittima e necessaria dell'architettura democratica.

Il passaggio da fasi sperimentali a forme istituzionali della partecipazione così intesa può avvenire mediante tre meccanismi principali: l'incentivazione, di tipo reputazionale o legata a finanziamenti da parte di istituzioni quali l'Unione europea, l'OCSE e il Consiglio d'Europa; l'incorporazione culturale, ossia la normalizzazione delle pratiche nei contesti istituzionali esistenti attraverso l'apprendimento tra pari; e la persuasione, che implica l'interiorizzazione normativa delle innovazioni democratiche come legittime e desiderabili (Pudar Drasko e Fiket 2026). Entro questa dinamica, il tessuto para-normativo si configura come componente fondamentale della trasformazione della governance europea – e nazionale – in senso partecipativo, attraverso lo strumento dell'orientamento più che dell'ancoraggio legale.

Quando la lente si restringe sull'Italia, il quadro appare coerente con una diffusione "a macchie", frutto di circolazioni transnazionali e sperimentazioni territoriali, che nel tempo hanno prodotto una famiglia di pratiche e dispositivi variabili per intensità e scala, spesso radicati nei governi locali e regionali (Allegretti 2011, 1). L'ancoraggio giuridico-costituzionale della partecipazione rappresenta il fondamento che consente l'emergere di istituti e procedure, raccomandando al contempo prudenza rispetto a irrigidimenti normativi precoci dato il carattere sperimentale e riflessivo di molte innovazioni partecipative (ivi, 7). Il dibattito scientifico sull'opportunità di istituzionalizzare resta strutturalmente bifronte: da un lato chiarezza, legittimità e stabilità, e dall'altro i rischi di ritualizzazione, controllo top-down e riproduzione di asimmetrie di potere (Brunazzo 2017, 841).

Istituzionalizzare la partecipazione pubblica non equivale qui a creare nuove arene generali e permanenti che riproducano i meccanismi della democrazia rappresentativa, ma a stabilizzare nel tempo l'uso di dispositivi partecipativi, attivati in relazione a determinate politiche, beni, interventi. La stabilità e la continuità riguardano dunque l'assetto complessivo – la presenza di norme, procedure e infrastrutture che mirano a rendere ordinario il coinvolgimento della cittadinanza – mentre le singole arene partecipative mantengono un carattere tematico e temporalmente circoscritto, in coerenza con la distinzione tra arene generaliste stabili e arene *single-issue* (Bobbio e Pomatto 2007, 2-3). La democrazia partecipativa istituzionalizzata non si configura in questo senso come surrogato della democrazia rappresentativa.

3. La metodologia

La mappatura[15] dei dispositivi considerati mira a individuare le tappe attraverso cui la partecipazione si è istituzionalizzata nel contesto italiano.

La ricognizione è finalizzata all'individuazione e alla sistematizzazione dei documenti – normativi e para-normativi – che hanno contribuito al processo di formalizzazione, istituzionalizzazione e infrastrutturazione della partecipazione pubblica con un approccio multilivello. Sono stati quindi presi in considerazione atti di *hard law* che introducono o disciplinano istituti giuridici di partecipazione, inclusi gli atti di pianificazione e programmazione formalmente approvati, e documenti di *soft law* con funzione di indirizzo, standardizzazione e consolidamento degli *habitus* collaborativi[16].

14 Come Unione europea, Consiglio d'Europa e UNECE.

15 Disponibile in appendice. La ricostruzione proposta non ha pretese di esaustività. Essa rappresenta una mappatura perfezionabile da molteplici punti di vista, sia rispetto all'ampliamento delle fonti considerate sia in relazione alla loro possibile riclassificazione. L'ordinamento cronologico dei dispositivi segue l'anno di approvazione o adozione, senza un ulteriore livello di dettaglio relativo a mese e giorno.
Tra le fonti utilizzate per l'individuazione preliminare degli atti regionali figura la ricognizione della Regione Emilia-Romagna (2016), successivamente verificata e integrata.

16 La selezione ha privilegiato gli atti che definiscono diritti, procedure o standard di coinvolgimento della cittadinanza nei processi di azione pubblica, che concorrono alla normalizzazione della partecipazione quale componente strutturale dell'azione amministrativa. Sono stati esclusi i documenti meramente descrittivi, le ricerche empiriche e i report privi di funzione regolativa o di standardizzazione.

Tale mappatura va intesa come una lettura parziale del fenomeno, poiché – naturalmente – l'istituzionalizzazione eccede la dimensione regolativa e si dispiega anche attraverso pratiche e configurazioni organizzative formali e informali, dinamiche di implementazione e processi di apprendimento istituzionale che, nel tempo, possono consolidare oppure indebolire quanto previsto sul piano normativo.

Nello specifico, sono stati inclusi gli atti: a) che disciplinano o promuovono diritti e forme di consultazione o partecipazione effettiva, intendendo con ciò il coinvolgimento – a diversa intensità – dei cittadini, singoli o organizzati, nei processi di elaborazione delle politiche pubbliche o nello svolgimento condiviso di attività di interesse generale; b) di livello nazionale, regionale, delle province autonome e delle città metropolitane, per ricostruire la stratificazione multilivello e le modalità di recepimento e adattamento delle cornici nazionali da parte degli enti territoriali; c) che configurano la partecipazione pubblica come politica (ad esempio le leggi organiche regionali) o che la inquadrano come insieme di strumenti di policy in senso procedurale (forme settoriali di partecipazione); d) identificabili come 'atti matrice', ossia testi che hanno funto da riferimento per successive normazioni o che hanno inaugurato cornici procedurali poi consolidate; e) nazionali di ratifica o recepimento di strumenti internazionali o europei, qualora la loro attuazione abbia effettivamente introdotto nell'ordinamento diritti o istituti partecipativi dotati di rilevanza giuridico-amministrativa, contribuendo alla domesticazione di standard sovranazionali entro la cultura amministrativa italiana.

Sono stati invece esclusi gli atti: f) che si limitano a prevedere principi di trasparenza o accesso alle informazioni (con eccezione della norma sulla partecipazione al procedimento e di alcune disposizioni regionali rilevanti ai fini della ricostruzione diacronica), nonché quelli riferiti esclusivamente a strumenti di democrazia diretta o annoverabili in altri domini democratici, o ancora a forme di partecipazione che non prevedano un'interazione strutturata tra cittadinanza e istituzioni pubbliche; g) di livello comunale, per ragioni di coerenza analitica e di spazio espositivo, salvo alcuni casi paradigmatici rilevanti per la traiettoria istituzionale; h) di recepimento o ridondanti rispetto a testi superiori, salvo quando tali atti introducano elementi partecipativi autonomi rispetto alla disciplina sovraordinata; i) prodotti dalla giurisprudenza, da autorità indipendenti o da altri organi di regolazione, pur riconoscendo che tali fonti hanno contribuito in modo significativo alla definizione dei contorni applicativi e all'evoluzione dei processi di istituzionalizzazione; l) internazionali non recepiti nell'ordinamento italiano o non tradotti in atti endogeni dotati di rilevanza partecipativa, ferma restando la loro influenza indiretta sulle trasformazioni normative e culturali del sistema italiano.

Sono stati presi in considerazione in totale 93 documenti. Per ciascun atto sono stati rilevati, in forma tabellare, il titolo, l'ambito di policy, la scala, la tipologia del dispositivo e la data di approvazione o pubblicazione.

Sulla base di tale ricostruzione, si propone a seguire una periodizzazione del percorso italiano di istituzionalizzazione della partecipazione pubblica, che si fonda sull'individuazione di discontinuità qualitative nel modo in cui la partecipazione è stata progressivamente incorporata nelle strutture istituzionali.

Alcune precisazioni

La scelta di concentrare la ricognizione a partire dagli anni Ottanta non implica – come anticipato – la negazione di precedenti tentativi di istituzionalizzazione della partecipazione che hanno interessato il sistema italiano, in particolare negli anni Settanta. Ad esempio, con l'istituzione degli organi collegiali nelle scuole e dei consigli di circoscrizione nei quartieri, la partecipazione della cittadinanza veniva "canalizzata" (Bobbio e Pomatto 2007, 2-3) entro organi che ricalcavano i meccanismi – e dunque i limiti – della democrazia rappresentativa, senza mettere in discussione la struttura dell'arena decisionale. Se da un lato tali risposte istituzionali possono essere lette come esperimenti di apertura alla partecipazione pubblica istituzionalizzata, dall'altro si è resa evidente, nel tempo, la progressiva ritualizzazione di quegli spazi e il riflusso della mobilitazione civica, con una formalizzazione che tende a normalizzare e attenuare la carica conflittuale, configurando la partecipazione più come rituale che come effettivo potere.

Per ragioni di delimitazione empirica, il presente contributo non ricostruisce sistematicamente tale stagione, ma assume come punto di avvio della periodizzazione la fase in cui la partecipazione entra nei testi normativi e para-normativi come componente esplicita dell'azione amministrativa a partire dagli anni Ottanta.

Al netto di ciò, nel contributo vengono esclusi gli organi generalisti e stabili che riproducono su scala ridotta la logica parlamentare. Diversamente, si è ritenuto opportuno considerare consulte e organismi tematici che: (i) operano in ambiti di policy specifici; (ii) svolgono funzioni consultive e propositive legate a piani, programmi o atti

settoriali; (iii) sono agganciati a processi strutturati (conferenze, consultazioni periodiche, valutazioni di servizi, ecc.); (iv) sono composti da attori organizzati o selezionati tramite procedure specifiche. Occorre precisare che le consulte tematiche occupano una posizione intermedia tra la rappresentanza organizzata e i dispositivi partecipativi procedurali propriamente detti, in quanto strutturano un canale stabile di consultazione tra cittadinanza e istituzioni senza sovrapporsi agli organi rappresentativi elettivi.

Sempre nell'ambito di un dialogo più strutturato, tre dimensioni hanno orientato la distinzione tra partecipazione di attori organizzati della società civile e concertazione classica: (i) l'apertura del perimetro degli attori coinvolti, che non si limita a sindacati, associazioni datoriali e ordini professionali; (ii) la presenza di una procedura esplicita accompagnata da regole di trasparenza su chi partecipa, quando e come; (iii) l'aggancio al processo di elaborazione delle politiche[17]. Quando una disposizione normativa si limita a regolare relazioni bilaterali tra governo e parti sociali non entra nel nucleo oggetto di mappatura.

4. La periodizzazione

Il processo italiano di istituzionalizzazione della partecipazione pubblica risulta profondamente condizionato dall'assetto decentralizzato dell'ordinamento e dalla distribuzione multilivello delle competenze amministrative e legislative, che hanno favorito una diffusione policentrica e non omogenea dei dispositivi. La riforma del Titolo V della Costituzione del 2001 ha rafforzato il principio di prossimità amministrativa e la valorizzazione dell'iniziativa civica, offrendo un ancoraggio normativo che ha reso più praticabili tanto la sperimentazione quanto la successiva stabilizzazione di pratiche partecipative e collaborative soprattutto alla scala territoriale.

Al livello nazionale, gli approcci più visibili si sono per lo più concentrati su dispositivi procedurali e di settore, nonché sulla cornice del governo aperto. Al livello territoriale, invece, l'istituzionalizzazione ha assunto più frequentemente la forma di leggi e politiche riconducibili sia al paradigma della democrazia partecipativa sia, in misura crescente, a quello dell'amministrazione condivisa, in coerenza sia con la capacità delle Regioni e dei Comuni di fungere da laboratori normativi, sia con la loro maggiore prossimità ai problemi pubblici e ai conflitti locali.

La divergenza tra i diversi livelli di governo riflette la frammentazione delle competenze e delle capacità amministrative, che orienta il livello centrale verso strumenti più standardizzabili e generalizzabili, mentre Regioni ed Enti locali dispongono di margini più ampi per costruire dispositivi contestuali e, talora, più ambiziosi. La traiettoria dell'amministrazione condivisa restituisce con evidenza questo esito policentrico, poiché dispositivi dapprima consolidati in ambito comunale possono essere successivamente sostenuti e amplificati da interventi regionali, come mostrano le relative normative di promozione dell'amministrazione condivisa dei beni comuni. Ne deriva un sistema nel quale la partecipazione tende a istituzionalizzarsi lungo filiere differenziate.

Ciascuna fase della seguente periodizzazione corrisponde a un diverso stadio di maturazione del rapporto tra cittadinanza e istituzioni, ricostruendo la transizione da una fase proto-partecipativa di proceduralismo garantista alla fase contemporanea, caratterizzata da un consolidamento multilivello e da un progressivo orientamento *soft-law-driven*.

Proceduralismo garantista e prime aperture settoriali (1980-1999)

Il primo periodo, il cui ancoraggio può essere simbolicamente ricondotto al principio enunciato dall'art. 3 comma 2 della Costituzione e la cui conclusione può essere collocata nel 1999, si configura come una fase proto-partecipativa, segnata da interventi ancora frammentari e privi di una compiuta sistematicità. La partecipazione emerge in ambiti specifici – sanità, ambiente, urbanistica – come dispositivo funzionale 'ad altro', e non ancora come dimensione autonoma di dialogo o collaborazione tra istituzioni e cittadinanza.

L'ingresso della partecipazione nel diritto amministrativo avviene in termini prettamente garantistici, quale standard di correttezza amministrativa. La domanda di controllabilità e correttezza dell'azione pubblica cresce nel contesto di crisi politico-istituzionale dei primi anni Novanta, contribuendo a ristrutturare il rapporto tra cittadini e istituzioni e a rafforzare l'attenzione per trasparenza e vincoli procedimentali (Di Mascio e Natalini 2024, 13). La legge n. 241/1990, nel codificare il diritto dei cittadini a partecipare al procedimento amministrativo, configura la partecipazione quale strumento volto ad assicurare la correttezza dell'azione pubblica: il cittadino vi è concepito come portatore di interessi legittimi giuridicamente qualificati, da tutelare

17 In fase di analisi e selezione dei riferimenti, questi criteri sono stati operazionalizzati interrogando principalmente tre dimensioni: chi partecipa (solo parti sociali organizzate o anche enti del Terzo settore, singoli cittadini ecc.), come partecipa (tavoli chiusi e prestabiliti o processi aperti o in cui sono esplicitate le modalità di selezione) e su cosa incide (accordi tra governo e parti sociali, progettazioni di piani e servizi, gestione di beni ecc.).

entro un processo regolato, non come soggetto attivo nella definizione dell'interesse generale. Le riforme amministrative della seconda metà degli anni Novanta – decentramento, semplificazione, riassetti organizzativi[18] – predispongono a loro volta condizioni favorevoli a una successiva differenziazione territoriale dell'offerta partecipativa, che diverrà più evidente nelle fasi seguenti.

Già in questo periodo, tuttavia, alcune Regioni anticipano per via settoriale una parziale apertura alla dimensione dialogica (come mostrano ad es. la L.R. Lazio n. 74/1991 e la L.R. Liguria n. 38/1998). Si tratta di interventi che prefigurano spazi ancora embrionali di coinvolgimento collettivo, pur restando inscritti entro logiche essenzialmente procedurali. Nel complesso, questa fase segna il passaggio da un modello amministrativo chiuso a una prima, limitata e spesso formale, apertura nei confronti della cittadinanza.

Sussidiarietà e welfare partecipativo (2000-2006)

Il periodo 2000-2006 rappresenta una discontinuità qualitativa, nella misura in cui la partecipazione entra nei processi di programmazione delle politiche sociali e acquista una dimensione più strutturale. La legge n. 328/2000 introduce forme organizzate di collaborazione tra enti locali e attori sociali, delineando un modello di welfare territoriale fondato sulla programmazione partecipata; attraverso i Piani di zona (art. 19), essa mira infatti a coinvolgere le risorse della comunità e a "responsabilizzare i cittadini nella programmazione e nella verifica dei servizi", rappresentando "un autentico *turning point*" (Burgalassi *et al*. 2021, 7) per lo sviluppo e l'innovazione dei servizi alla persona. È in questo contesto che la partecipazione entra nel lessico delle politiche pubbliche, assumendo la funzione di metodo e non più soltanto quella di garanzia procedimentale.

In parallelo, il Testo unico degli enti locali (D.Lgs. n. 267/2000) consolida il ruolo dei Comuni, che "promuovono organismi di partecipazione popolare all'amministrazione locale", configurandoli come sedi privilegiate di relazione con la cittadinanza. Successivamente il D.P.C.M. 30 marzo 2001, di attuazione della legge n. 328/2000, introduce per la prima volta nel lessico giuridico-amministrativo il termine "coprogettazione", per individuare insieme ai soggetti del Terzo settore progetti sperimentali e innovativi diretti ad affrontare specifiche problematiche sociali.

La riforma del Titolo V della Costituzione del 2001 segna un ulteriore punto di svolta: l'art. 118 comma 4 riconosce costituzionalmente e promuove l'autonoma iniziativa dei cittadini per lo svolgimento di attività di interesse generale. La sussidiarietà costituzionale assume così portata al tempo stesso garantista e programmatoria, diventando il riferimento teorico e normativo di successive innovazioni legislative. Si tratta di un passaggio fondamentale, poiché per la prima volta la collaborazione civica viene inscritta nel quadro costituzionale come dimensione propria dell'azione pubblica. L'Umbria (L.R. n. 16/2006) è la prima Regione a recepire nel proprio ordinamento il paradigma collaborativo fondato sulla sussidiarietà, aprendo la strada alle successive iniziative regionali[19].

Sul fronte ambientale, il decreto legislativo n. 152/2006 sistematizza al livello nazionale le procedure partecipative nelle valutazioni ambientali (VAS, VIA), introducendo fasi obbligatorie che formalizzano un dialogo istituzionalizzato con cittadini e portatori di interesse. L'evoluzione di questa fase è stata dunque favorita sia dalla crescente centralità di reti e attori sociali territoriali nella programmazione – coerente con l'idea che i problemi pubblici richiedano un coordinamento multi-attore – sia dalla costituzionalizzazione del principio di sussidiarietà. La partecipazione evolve così da strumento di tutela a tecnica di policy, fondata su dialogo, partenariato e corresponsabilità nella definizione e nell'attuazione degli interventi, con esiti differenziati tra territori in funzione delle capacità amministrative e delle configurazioni istituzionali locali.

Sperimentazione istituzionale e modelli regionali apripista (2007-2013)

Tra il 2007 e il 2013 si apre una fase di sperimentazione istituzionale caratterizzata dall'emergere di modelli regionali che concepiscono la partecipazione come politica autonoma. Le regioni – che agiscono in qualità di *policy entrepreneurs* in un assetto decentrato – operano come laboratori istituzionali, intercettando una domanda di innovazione amministrativa e traducendola in strumenti replicabili. Fattori abilitanti di tale dinamica sono, da un lato, l'autonomia normativa regionale e, dall'altro, la presenza di comunità professionali della partecipazione, spesso in connessione con coalizioni civiche territoriali.

La L.R. Toscana n. 69/2007[20] viene considerata

18 Nel quadro della c.d. riforma Bassanini.

19 Cfr. Giglioni F. (2016), Legislazione regionale in materia di sussidiarietà orizzontale, *Labsus*, https://www.labsus.org/2016/12/legislazione-regionale-in-materia-di-sussidiarieta-orizzontale (consultato il 05/10/2025).

20 Si veda, ad esempio, Floridia 2008.

il primo esempio di disciplina organica della partecipazione in Italia, ponendo le basi di un modello che verrà successivamente ripreso e adattato da altre amministrazioni regionali. Essa istituisce meccanismi procedurali per l'attivazione di processi partecipativi, prevede il sostegno finanziario ai percorsi deliberativi locali e introduce un'autorità indipendente con funzioni di garanzia e di monitoraggio. Esperienze analoghe si sviluppano negli anni successivi, come dimostra la successiva L.R. Emilia-Romagna n. 3/2010. La partecipazione cessa in tal modo di configurarsi come mero adempimento normativo e diviene una scelta politica deliberata da parte di Regioni e Province autonome.

Accanto a queste esperienze legislative regionali, leggibili come modelli di *policy design* partecipativo, si consolidano iniziative e strumenti di carattere settoriale che ampliano la diffusione di pratiche deliberative e collaborative, come la Carta nazionale dei Contratti di fiume[21]. Nella medesima fase si rafforza, inoltre, un canale di istituzionalizzazione 'morbida' legato a cornici internazionali di governo aperto: l'adesione dell'Italia all'Open Government Partnership, a partire dal 2011, introduce difatti una logica di *accountability* multilivello e favorisce la formazione di reti professionali intorno al paradigma del governo aperto, alimentando processi di standardizzazione e di costruzione di reti e routine amministrative (quali piani d'azione, consultazioni, definizione della governance e dei ruoli, e così via).

La partecipazione assume progressivamente una valenza più chiaramente politica, connessa all'innovazione amministrativa, allo sviluppo territoriale e al rafforzamento della fiducia. La questione diventa come rendere la partecipazione attivabile e sostenibile entro dispositivi e infrastrutture differenziati per scala e ambito, con effetti in termini di apprendimento istituzionale e diffusione.

Istituzionalizzazione multilivello e costruzione delle infrastrutture (2014-2020)

Questa fase è probabilmente la più densa e composita dell'intero processo di istituzionalizzazione, poiché in essa convergono traiettorie diverse. Il periodo 2014-2020 è infatti caratterizzato da un'espansione multilivello della partecipazione, che si consolida come componente della governance attraverso il proseguimento dell'adozione di leggi regionali dedicate e la creazione e il rafforzamento di infrastrutture. Le dinamiche di coinvolgimento civico si articolano progressivamente lungo l'asse verticale che va dal livello nazionale a quello locale, generando una rete di norme, strumenti, strutture e meccanismi di coordinamento.

Sul piano locale, il Regolamento di Bologna del 2014 inaugura un modello di amministrazione condivisa per la cura dei beni comuni che, pur non esaurendo la varietà delle esperienze italiane sviluppatesi, rappresenta il punto di avvio di una trasformazione poi replicata da numerosi Enti locali – così spingendo l'istituzionalizzazione lungo la scala comunale e affiancando alla partecipazione consultiva forme di co-gestione, soprattutto attraverso lo strumento del patto di collaborazione[22]. Tale traiettoria urbana sposta l'attenzione verso dispositivi di corresponsabilità operativa, rendendo più ordinaria la collaborazione tra istituzioni e cittadinanza nelle pratiche di cura e rigenerazione e trovando un corrispondente sostegno, al livello regionale, nelle leggi di promozione dell'amministrazione condivisa dei beni comuni (come la L.R. Lazio n. 10/2019).

Sul piano nazionale, il D.Lgs. n. 50/2016 introduce – e in alcuni casi rende obbligatorio – il dibattito pubblico per le grandi opere, concorrendo a istituzionalizzare forme di partecipazione deliberativa nella pianificazione infrastrutturale. Contestualmente, documenti di *soft law* nazionale quali le Linee guida sulla consultazione (2017, 2020) e la Carta della partecipazione (2014, 2024) forniscono un *framework* metodologico condiviso e svolgono una funzione di standardizzazione dei diritti e delle pratiche, orientando l'uniformazione di linguaggi e routine amministrative.

All'interno di questa fase può essere individuata un'ulteriore cesura tra il 2016 e il 2017, segnata dalla riforma e dall'adozione del Codice del Terzo settore (D.Lgs. n. 117/2017): tale intervento affianca e in parte ricompone le sperimentazioni maturate al livello territoriale entro una cornice normativa nazionale, in coerenza con il più generale riordino del settore. La L.R. Toscana n. 65/2020 rappresenta uno dei primi esempi di recepimento organico della riforma in ambito regionale, segnando il passaggio verso una fase di consolidamento e di diffusione del paradigma collaborativo regolamentato.

Ne risulta una struttura partecipativa articolata, policentrica e multilivello, alimentata da un intreccio di *hard* e *soft law* e sorretta da un progressivo processo di infrastrutturazione.

21 Per la definizione dello strumento cfr. D.Lgs. n. 152/2006 e successive modificazioni e integrazioni.

22 Per la definizione dello strumento cfr. Comune di Bologna (2014), Regolamento sulla collaborazione tra cittadini e amministrazione per la cura e la rigenerazione dei beni comuni urbani, https://www.comune.bologna.it/myportal/C_A944/api/content/download?id=5eb41e5ad343fb00803df458

Consolidamento procedurale e integrazione trasversale (2021-oggi)

L'ultima fase, che si apre nel 2021 e giunge fino a oggi, è contraddistinta da un processo di consolidamento procedurale e di progressiva integrazione trasversale della partecipazione. Il coinvolgimento della cittadinanza si mostra estendibile a una pluralità di settori dell'azione amministrativa e si afferma effettivamente in campi 'nuovi' – parità di genere, energia rinnovabile, infanzia e invecchiamento – attraverso documenti strategici e linee guida che ne integrano le istanze in più processi del policy-making. Questi ambiti evidenziano una crescente ibridazione tra partecipazione, inclusione e innovazione sociale, che consolida il ruolo dei cittadini e delle organizzazioni civiche quali co-produttori di valore pubblico, con effetti di professionalizzazione e di stabilizzazione delle aspettative tra attori; questo tipo di partecipazione tende a divenire uno strumento riconosciuto e legittimato tra i *practitioners*, pur continuando a presentare differenze territoriali legate tanto alle capacità amministrative quanto alle scelte politiche.

Una spinta decisiva deriva da dispositivi quali il PNRR e il D.M. n. 72/2021, che orientano l'uso degli istituti previsti dagli artt. 55-57 del Codice del Terzo settore e contribuiscono a rendere co-programmazione e co-progettazione strumenti ordinari di collaborazione tra enti del Terzo settore e amministrazioni pubbliche, sottraendoli alla dimensione sperimentale e aumentandone la ripetibilità. Le Regioni e le Città metropolitane incorporano la partecipazione nelle strategie territoriali (ad es. nella Strategia Valle d'Aosta 2030 e nel PUMS della Città metropolitana di Roma Capitale), mentre le Linee guida nazionali per la partecipazione pubblica e quelle sul Piano Integrato di Attività e Organizzazione (PIAO) si propongono di coordinare questo ecosistema e di ridurre la frammentazione esistente, ancorando il coinvolgimento civico anche agli strumenti di programmazione e controllo interno e rafforzando la tendenza a costruire meccanismi di standardizzazione a scala nazionale.

L'istituzionalizzazione della partecipazione si accompagna, tuttavia, anche a modifiche nelle previsioni normative che ne rivelano la natura contesa, dinamica e reversibile. La parabola del dibattito pubblico – introdotto nel 2016 e poi riformato nel 2023 – offre evidenza di questa dinamica, poiché dopo una fase di rafforzamento del radicamento procedurale l'assetto successivo può ridurne la portata e la capacità di incidenza, fino a ricondurlo a funzioni più prossime alla mera informazione e all'accesso documentale[23].

Conclusioni

L'analisi condotta mostra come la partecipazione pubblica si configuri come un campo stratificato, la cui intelligibilità richiede una preliminare ricomposizione teorica dei paradigmi di riferimento. La prima parte del contributo ha messo in luce come l'articolazione dei rapporti tra democrazia partecipativa, governo aperto e amministrazione condivisa consenta di ricondurre tale pluralità entro una griglia concettuale capace di rendere conto della diversità delle pratiche.

La nozione di istituzionalizzazione designa il processo mediante il quale pratiche inizialmente sperimentali, situate e dipendenti da contingenze locali, vengono incorporate in schemi di azione regolati e riproducibili. La mappatura dei dispositivi di *hard* e *soft law* adottati ai diversi livelli dell'ordinamento italiano evidenzia come tale processo sia avvenuto secondo logiche non lineari. Le fasi individuate mostrano come la partecipazione sia passata da un uso prevalentemente strumentale e garantista a una progressiva configurazione come politica pubblica autonoma, fino alla recente emersione di un'infrastruttura multilivello e *soft-law-driven*.

Tale evoluzione presenta tuttavia effetti ambivalenti. La standardizzazione, utile ai fini di stabilità e trasparenza, può tradursi in un irrigidimento che limita la capacità adattiva dei processi e ne riduce la responsività rispetto ai contesti sociali. Analogamente, la spersonalizzazione dei dispositivi può allontanare le pratiche dal loro originario carattere sperimentale e trasformativo, producendo una burocratizzazione della collaborazione civica e rischiando di degenerare in un rituale tecnocratico privo di incidenza. Ciononostante, tali spazi possono nel tempo operare come dispositivi di apprendimento istituzionale, favorendo la formazione di un *habitus* amministrativo e, potenzialmente, trasformandosi in arene di redistribuzione simbolica e sostanziale delle risorse.

La periodizzazione proposta richiederebbe, per ciascuna fase, un approfondimento più esteso, anche

23 Ravazzi (2026), con riferimento al caso italiano, osserva che l'adozione di un quadro normativo può rafforzare l'*embeddedness* della deliberazione, aumentando la 'serietà' con cui attori istituzionali e tecnici gestiscono l'interazione con le comunità e accrescendo l'autorevolezza dei professionisti della partecipazione, con ricadute sulla capacità di incidere sugli esiti progettuali. La stessa proceduralizzazione può però alimentare processi di routinizzazione, spostando l'attenzione verso l'adempimento e riducendo flessibilità e adattamento al contesto, con effetti ambivalenti sull'integrazione della deliberazione nel sistema democratico.

alla luce del numero complessivo di documenti censiti e sistematizzati nell'Appendice. Al contempo è opportuno ribadire il carattere inevitabilmente parziale della ricostruzione, poiché i dispositivi formali e istituzionali rappresentano soltanto la 'punta dell'iceberg' di un fenomeno che eccede la dimensione normativa. Proprio per questa ragione, l'ordinamento sistematico delle tracce regolative mantiene una propria utilità analitica, poiché consente di ricostruire una sequenza ordinata e di metterla in relazione con i fattori politici e sociali che ne condizionano l'adozione, la stabilizzazione e gli effetti diretti e indiretti. In una prospettiva di ricerca ulteriormente sviluppabile, risulterebbe auspicabile connettere la traiettoria italiana con quella istituzionale internazionale, al fine di osservare con maggiore precisione la circolazione di politiche e standard – specie nello spazio europeo – e le modalità con cui tali cornici vengono recepite e rielaborate nei contesti nazionali.

Nei termini qui ricostruiti, la partecipazione pubblica si configura come un insieme articolato di pratiche e dispositivi che opera a un duplice livello. Sul piano sistemico, essa contribuisce alla ridefinizione del rapporto tra cittadinanza e amministrazioni pubbliche, in cui queste ultime affiancano progressivamente alla funzione di regolazione verticale il ruolo di nodi di reti partecipative. Sul piano operativo, e in prospettiva trasversale, introduce modalità procedurali che orientano la produzione delle politiche lungo l'intero processo di elaborazione, creando spazi regolati in cui cittadini e istituzioni interagiscono secondo diversi livelli di intensità.

Il contributo mostra, infine, come la partecipazione pubblica rappresenti un campo di policy emergente la cui istituzionalizzazione, lungi dall'essere un mero processo tecnico-amministrativo, riflette e al contempo orienta la trasformazione dei paradigmi democratici contemporanei.

Appendice

Dispositivi rilevanti per l'istituzionalizzazione della partecipazione pubblica in Italia (1983-2025)

N.	Titolo	Ambito di policy	Scala	Tipo	Anno di approvazione/ pubblicazione
1	L.R. Lazio n. 58/1983	Unità sanitarie locali	Regionale	*Hard law*	1983
2	L. n. 439/1989	Ratifica Carta europea dell'autonomia locale	Nazionale	*Hard law*	1989
3	L. n. 241/1990	Procedimento amministrativo	Nazionale	*Hard law*	1990
4	L.R. Lazio n. 74/1991	Tutela ambientale	Regionale	*Hard law*	1991
5	L.R. Lazio n. 35/1992	Attività amministrativa	Regionale	*Hard law*	1992
6	L.P. Trento n. 4/1996	Programmazione di sviluppo	Provinciale	*Hard law*	1996
7	L.R. Liguria n. 38/1998[24]	Valutazione impatto ambientale	Regionale	*Hard law*	1998
8	L.R. Lazio n. 30/1998	Trasporto pubblico	Regionale	*Hard law*	1998
9	L.R. Lazio n. 38/1999	Governo del territorio	Regionale	*Hard law*	1999
10	L. n. 328/2000	Politiche sociali	Nazionale	*Hard law*	2000
11	D.Lgs. n. 267/2000[25]	Enti locali	Nazionale	*Hard law*	2000

24 Non riportate le ulteriori leggi regionali in materia di VIA (come la L.R. Emilia-Romagna n. 9/1999) che precedono la sistematizzazione nazionale operata dal D.Lgs. n. 152/2006.

25 Non riportate le leggi regionali che, pur anticipando il modello di partecipazione locale poi generalizzato dal TUEL, restano circoscritte a singoli ordinamenti, come la L.R. Valle d'Aosta n. 54/1998.

12	L.R. Emilia-Romagna n. 20/2000[26]	Governo del territorio	Regionale	*Hard law*	2000
13	D.P.C.M. 30 marzo 2001	Welfare, Terzo settore	Nazionale	*Hard law*	2001
14	L.R. Lazio n. 25/2001	Programmazione, bilancio	Regionale	*Hard law*	2001
15	L. Cost. n. 3/2001 (art. 118 comma. 4)	Sussidiarietà costituzionale	Nazionale	*Hard law*	2001
16	L. n. 108/2001	Ratifica Aarhus	Nazionale	*Hard law*	2001
17	L.R. Lombardia n. 26/2003	Servizi locali, risorse idriche	Regionale	*Hard law*	2003
18	D.Lgs. n. 42/2004	Beni culturali, paesaggio	Nazionale	*Hard law*	2004
19	D.Lgs. n. 82/2005	Amministrazione digitale	Nazionale	*Hard law*	2005
20	L.R. Lazio n. 4/2006[27]	Finanziaria	Regionale	*Hard law*	2006
21	L. n. 14/2006	Ratifica Convenzione paesaggio	Nazionale	*Hard law*	2006
22	Direttiva del Ministro per le riforme e le innovazioni nella Pubblica amministrazione per una Pubblica amministrazione di qualità	Pubblica amministrazione	Nazionale	*Soft law*	2006
23	L.R. Umbria n. 16/2006[28]	Sussidiarietà costituzionale	Regionale	*Hard law*	2006
24	D.Lgs. n. 152/2006[29]	Ambiente, VAS/VIA	Nazionale	*Hard law*	2006
25	L.P. Trento n. 3/2006 (come modificata dalla L.P. 12/2014)	Governo dell'autonomia	Provinciale	*Hard law*	2006
26	L.R. Toscana n. 69/2007[30]	Partecipazione	Regionale	*Hard law*	2007
27	L.R. Lazio n. 20/2007	Partecipazione giovanile	Regionale	*Hard law*	2007
28	L.P. Trento n. 13/2007	Politiche sociali	Provinciale	*Hard law*	2007
29	L.R. Liguria n. 6/2009	Minori e giovani	Regionale	*Hard law*	2009
30	D.Lgs. n. 150/2009	Pubblica amministrazione	Nazionale	*Hard law*	2009
31	L.R. Emilia-Romagna n. 3/2010[31]	Partecipazione	Regionale	*Hard law*	2010
32	L.R. Umbria n. 14/2010	Partecipazione	Regionale	*Hard law*	2010
33	L.P. Trento n. 16/2010	Salute	Provinciale	*Hard law*	2010
34	L.R. Abruzzo n. 26/2010[32]	Qualità della normazione	Regionale	*Hard law*	2010
35	Carta Nazionale dei Contratti di fiume[33]	Gestione risorse idriche, ambiente	Nazionale	*Soft law*	2010
36	L.R. Emilia-Romagna n. 23/2011	Servizi pubblici ambientali	Regionale	*Hard law*	2011
37	Linee guida sulla definizione di criteri e modelli per la partecipazione del Terzo settore alla determinazione delle politiche pubbliche a livello locale	Terzo settore	Nazionale	*Soft law*	2011

26 Non riportate le ulteriori leggi regionali sul governo del territorio che prevedono forme analoghe di partecipazione, come la L.R. Calabria n. 19/2002 e la L.R. Lombardia n. 12/2005.

27 Non riportata la successiva legge finanziaria L.R. Lazio n. 26/2007.

28 Non riportate le successive leggi di applicazione del principio di sussidiarietà costituzionale: L.R. Campania n. 12/2011, L.R. Calabria n. 29/2012, L.R. Liguria n. 13/2015, L.R. Piemonte n. 10/2016.

29 Non riportate le successive leggi regionali di attuazione della VAS/VIA, come la L.R. Liguria n. 32/2012.

30 Abrogata.

31 Abrogata.

32 Non riportate le ulteriori leggi regionali riconducibili alla disciplina della qualità della regolazione e della semplificazione amministrativa, come la L.R. Liguria n. 13/2011, la L.R. Marche n. 3/2015 e la L.R. Campania n. 11/2015.

33 Non riportati gli altri documenti di formalizzazione relativi a questo strumento: *Coinvolgimento e partecipazione dei portatori d'interesse – Approcci, metodi e strumenti per i Contratti di Fiume, L'approccio win-win nei Contratti di fiume*.

38	1° Piano d'Azione nazionale per l'Open Government Partnership	Governo aperto	Nazionale	*Soft law*	2012
39	Comunicare partecipazione. Strategia comunicativa integrata a supporto dei processi inclusivi	Comunicazione istituzionale, partecipazione	Regionale	*Soft law*	2012
40	L.R. Lombardia n. 1/2012	Procedimento amministrativo	Regionale	*Hard law*	2012
41	L.R. FVG n. 5/2012	Partecipazione giovanile	Regionale	*Hard law*	2012
42	L.R. FVG n. 23/2012	Volontariato, associazionismo	Regionale	*Hard law*	2012
43	L.P. Trento n. 5/2013[34]	Controllo sull'attuazione e valutazione	Provinciale	*Hard law*	2013
44	L.R. Toscana n. 46/2013	Partecipazione	Regionale	*Hard law*	2013
45	L.R. Lazio n. 2/2014	Protezione civile	Regionale	*Hard law*	2014
46	L.R. Lazio n. 5/2014	Governance delle acque	Regionale	*Hard law*	2014
47	L.R. Sicilia n. 5/2014[35]	Bilancio partecipativo	Regionale	*Hard law*	2014
48	Regolamento sulla collaborazione tra cittadini e amministrazione per la cura e la rigenerazione dei beni Comuni urbani - Comune di Bologna[36]	Amministrazione condivisa	Comunale	*Hard law*	2014
49	L.R. Toscana n. 65/2014[37]	Governo del territorio	Regionale	*Hard law*	2014
50	Accordo di Partenariato 2014-2020	Politiche di coesione, partenariato	Nazionale	*Soft law*	2014
51	Carta della Partecipazione	Partecipazione	Nazionale[38]	*Soft law*	2014[39]
52	L.R. Lazio n. 1/2015	Politiche internazionali	Regionale	*Hard law*	2015
53	L.R. Lazio n. 5/2015	Istruzione e formazione	Regionale	*Hard law*	2015
54	L.R. FVG n. 31/2015	Immigrazione, inclusione sociale	Regionale	*Hard law*	2015
55	L.R. Lombardia n. 36/2015	Cooperazione	Regionale	*Hard law*	2015
56	L.P. Trento n. 1/2016	Consulta per lo Statuto speciale	Provinciale	*Hard law*	2016
57	L.R. Lazio n. 11/2016	Interventi e servizi sociali	Regionale	*Hard law*	2016
58	L.R. Lazio n. 7/2016	Valutazione	Regionale	*Hard law*	2016
59	Regolamento per la disciplina del Bilancio Partecipativo - Comune di Bologna	Bilancio partecipativo	Comunale	*Hard law*	2016
60	D.Lgs. n. 50/2016, art. 22 e ss. (poi sostituito dal D.Lgs. n. 36/2023)	Dibattito pubblico	Nazionale	*Hard law*	2016
61	Linee guida sulla consultazione pubblica	Consultazione	Nazionale	*Soft law*	2017[40]

34 Non riportate le ulteriori leggi regionali in materia di controllo sull'attuazione delle leggi e di valutazione degli effetti delle politiche, come la L.R. Lombardia n. 20/2017.

35 Non riportate le successive leggi sul bilancio partecipativo.

36 Non riportati i successivi regolamenti adottati da regioni, comuni, città metropolitane, province, unioni di comuni e comunità montane.

37 Non riportate ulteriori discipline sul governo del territorio, come la L.P. Trento n. 15/2015.

38 Per i documenti non riconducibili a un'unica autorità o la cui portata territoriale non coincida con il relativo livello istituzionale, la scala indica il target territoriale d'azione, inteso come ambito di diffusione, coordinamento, recepimento o attuazione.

39 Aggiornata nel 2024.

40 Aggiornate nel 2020.

62	D.Lgs. n. 117/2017	Terzo settore	Nazionale	*Hard law*	2017
63	L.P. Trento n. 6/2017	Mobilità sostenibile	Provinciale	*Hard law*	2017
64	L.R. Puglia n. 28/2017	Partecipazione	Regionale	*Hard law*	2017
65	Il Bilancio di genere nell'approccio ben-essere BGbe - Linee guida per i Comuni[41]	Parità di genere	Regionale	*Soft law*	2018
66	L.P. Bolzano n. 22/2018	Partecipazione	Provinciale	*Hard law*	2018
67	L.R. Sardegna n. 22/2018	Politiche linguistiche	Regionale	*Hard law*	2018
68	L.R. Emilia-Romagna n. 15/2018	Partecipazione	Regionale	*Hard law*	2018
69	Linee guida per la disabilità e l'inclusione sociale negli interventi di cooperazione	Cooperazione internazionale, disabilità	Nazionale	*Soft law*	2018
70	D.P.C.M. n. 76/2018[42]	Dibattito pubblico	Nazionale	*Hard law*	2018
71	L.R. Lazio n. 10/2019[43]	Amministrazione condivisa	Regionale	*Hard law*	2019
72	Linee guida sulla valutazione partecipativa nelle amministrazioni pubbliche	Valutazione della performance	Nazionale	*Soft law*	2019
73	Linee guida per la realizzazione di sistemi di valutazione dell'impatto sociale delle attività svolte dagli enti del Terzo settore (D.M. 23 luglio 2019)	Terzo settore	Nazionale	*Soft law*	2019
74	Carta di intenti sulla trasparenza e partecipazione nelle valutazioni ambientali	Valutazioni ambientali	Nazionale	*Soft law*	2019
75	Anello Verde - Schema di Assetto Generale - Roma Capitale	Rigenerazione urbana, mobilità sostenibile	Comunale	*Soft law*	2020
76	L.R. Marche n. 31/2020	Partecipazione	Regionale	*Hard law*	2020
77	L.R. Toscana n. 65/2020[44]	Terzo settore	Regionale	*Hard law*	2020
78	Linee guida sull'uguaglianza di genere e l'empowerment di donne, ragazze e bambine (2020-2024)	Parità di genere	Nazionale/ Internazionale[45]	*Soft law*	2020
79	L. n. 133/2020	Ratifica Convenzione Faro	Nazionale	*Hard law*	2020
80	Piano Nazionale di Ripresa e Resilienza[46]	Governance collaborativa	Nazionale	*Hard law*	2021
81	Piano Strategico Nazionale sulla Violenza Maschile contro le Donne 2021-2023	Parità di genere, contrasto alla violenza	Nazionale	*Soft law*	2021
82	Linee guida sul rapporto tra pubbliche amministrazioni ed enti del Terzo settore (D.M. n. 72/2021)	Terzo settore	Nazionale	*Soft law*	2021

41 Non riportata la L.R. Emilia-Romagna n. 6/2014.

42 Non riportato il documento *Raccomandazione n. 1 – Linee guida sul* di*battito pubblico*.

43 Non riportate le successive leggi regionali riconducibili al filone dell'amministrazione condivisa: L.R. Toscana n. 71/2020, L.R. Umbria n. 2/2023, L.R. Emilia-Romagna n. 3/2023, L.R. Marche n. 23/2025, L.R. Puglia n. 11/2025.

44 Non riportate le successive leggi regionali in materia di Terzo settore, in particolare: L.R. Molise n. 21/2022, L.R. Piemonte n. 7/2024, L.P. Bolzano n. 7/2025.

45 Cfr. nota 38.

46 Non riportati gli atti normativi e amministrativi adottati in attuazione del PNRR, salvo quelli dotati di autonoma rilevanza ai fini della ricostruzione.

83	Strategia di Sviluppo Sostenibile della Valle d'Aosta 2030	Sviluppo sostenibile	Regionale	*Soft law*	2021
84	D.Lgs. n. 199/2021	Energia rinnovabile	Nazionale	*Hard law*	2021
85	L.R. Lazio n. 16/2021	Invecchiamento attivo	Regionale	*Hard law*	2021
86	Linee guida per la partecipazione di bambine e bambini e ragazze e ragazzi	Infanzia e adolescenza	Nazionale	*Soft law*	2022
87	Il ruolo della partecipazione e dello stakeholder engagement nella definizione e attuazione delle politiche per lo sviluppo sostenibile	Sviluppo sostenibile, Agenda 2030	Nazionale	*Soft law*	2023
88	Agire per ciò che più conta - Vademecum per la promozione della salute	Salute	Regionale	*Soft law*	2023
89	Linee Guida per il funzionamento e la gestione dell'Assemblea Permanente dei Cittadini sul Clima - Comune di Milano	Ambiente	Comunale	*Soft law*	2024
90	Ragazze, ragazzi e adulti nei processi partecipativi. Pratiche e strategie	Partecipazione giovanile	Nazionale	*Soft law*	2024
91	Piano Urbano della Mobilità Sostenibile (PUMS) della Città metropolitana di Roma Capitale	Mobilità	Metropolitana	*Hard law*	2024
92	Linee guida 2025 sul Piano Integrato di Attività e di Organizzazione (PIAO)	Programmazione e governance	Nazionale	*Soft law*	2025
93	Linee guida per la partecipazione pubblica	Partecipazione	Nazionale	*Soft law*	2025

Bibliografia

Allegretti U. (2011), La democrazia partecipativa in Italia e in Europa, *Rivista AIC*, n.1, pp.1-12

Angeli A. (2025), Le linee guida europee in materia di democrazia deliberativa e inclusione dei gruppi vulnerabili, *Nomos. Le attualità nel diritto*, n.1, pp.1-25

Ansell C., Gash A. (2008), Collaborative governance in theory and practice, *Journal of Public Administration Research and Theory*, 18, n.4, pp.543-571

Arena G. (2022), Un approccio sistemico all'amministrazione condivisa, in Arena G., Bombardelli M. (a cura di), *L'amministrazione condivisa*, Quaderni della Facoltà di Giurisprudenza n.62, Trento, Università degli Studi di Trento, pp.1-13

Arena G. (2004), *Un nuovo modo di amministrare*, Relazione tenuta alla Convenzione Nazionale della Sussidiarietà, Roma, 12 marzo 2004

Bevir M. (2010), Rethinking governmentality: Towards genealogies of governance, *European Journal of Social Theory*, 13, n.4, pp.423-441

Bifulco R. (2009), Democrazia partecipativa e democrazia deliberativa, *Relazione al Convegno "La democrazia partecipativa in Italia e in Europa: esperienze e prospettive",* Firenze, 2-3 aprile 2009, Astrid

Bobbio L., Pomatto G. (2007), Modelli di coinvolgimento dei cittadini nelle scelte pubbliche, *Rapporto per la Provincia Autonoma di Trento*

Bobbio L., Pomatto G., Ravazzi S. (2017), *Le politiche pubbliche. Problemi, soluzioni, incertezze, conflitti*, Milano, Mondadori Università

Boussaguet L. (2016), Participatory mechanisms as symbolic policy instruments?, *Comparative European Politics*, 14, n.1, pp.107-124

Brunazzo M. (2017), Istituzionalizzare la partecipazione? Le leggi sulla partecipazione in Italia, *Istituzioni del Federalismo*, n.3, pp.837-864

Burgalassi M., Galeone P., Pacini L. (a cura di) (2021), *I vent'anni della legge 328 del 2000 nella Penisola. Le trasformazioni del welfare locale*, Roma, Roma TrE-Press

D'Albergo E. (2014), *Sociologia della politica. Attori, strutture, interessi, idee*, Roma, Carocci

Di Mascio F., Natalini A. (2024), Le teorie del cambiamento amministrativo tra paradigmi e modelli. Il caso delle riforme amministrative in Italia, *Rivista Trimestrale di Scienza dell'Amministrazione*, n.1, pp.1-27

Floridia A. (2008), Democrazia deliberativa e processi decisionali: la legge della Regione Toscana sulla partecipazione, *Stato e Mercato*, n.82, pp.83-110

Fung A., Wright E. O. (eds.) (2003), *Deepening Democracy. Institutional Innovations in Empowered Participatory Governance*, London, Verso

Gaventa J. (2006), Finding the spaces for change: A power analysis, *IDS Bulletin*, 37, n.6, pp.23-33

Laino V. (2025), *Al di là del tavolo. Processi partecipativi e co-progettazione: modelli e strumenti*, Roma, Sapienza Università Editrice

Lewanski R. (2016), *La prossima democrazia. Dialogo – deliberazione – decisione*, s.l., Lulu

Mahoney J., Thelen K. (eds.) (2010), *Explaining Institutional Change. Ambiguity, Agency, and Power*, Cambridge, Cambridge University Press

Meijer A. J., Curtin D., Hillebrandt M. (2012), Open Government: Connecting vision and voice, *International Review of Administrative Sciences*, 78, n.1, pp.10-29

Michels A., Binnema H. (2019), Assessing the Impact of Deliberative Democratic Initiatives at the Local Level: A Framework for Analysis, *Administration & Society*, 51, n.5, pp.749-769

Moini G. (2012), *Teoria critica della partecipazione. Un approccio sociologico*, Milano, Franco Angeli

Moro G. (2022), Cittadini, dibattiti pubblici e legittimazione procedurale. Un early assessment dei dibattiti realizzati nel contesto del PNRR, *Studi Politici*, n.1, pp.147-164

Moro G. (2009), Partecipare a cosa? Per una riconsiderazione del nesso tra democrazia partecipativa e attivismo organizzato dei cittadini in Italia e in Europa, *Paper presentato al Convegno SISP,* Roma, LUISS, 18 settembre 2009

Newig J., Gaube V., Berkhoff K., Kaldrack K., Kastens B., Lutz J., Schlußmeier B., Adensam H., Haberl H. (2008), The role of formalisation, participation and context in the success of public involvement mechanisms in resource management, *Systemic Practice and Action Research*, 21, pp.423-441

OECD (2023), *Open Government for Stronger Democracies: A Global Assessment,* Paris, OECD Publishing

OECD (2016), *Open Government: The Global Context and the Way Forward*, Paris, OECD Publishing

Open Government Partnership (2025), *Co-creation and the Development of Action Plans,* Washington, DC, Open Government Partnership

Pezzana P. (2023), L'amministrazione condivisa: perseguire l'interesse generale tra processi e procedimento, *Impresa Sociale*, n.4, pp.29-35

Pudar Drasko G., Fiket I. (2026), Institutionalization of Participatory Democratic Innovations. Understanding the Roles of Established and Emerging Actors, *Politics and Governance*, 14, art. 10740

Randma-Liiv T. (2023), Adoption is not enough: Institutionalization of e-participation initiatives, *Public Policy and Administration*, 38, n.3, pp.329-351

Ravazzi S. (2026), The Legal Institutionalisation of Public Deliberation and the Embeddedness in the Democratic System. The Italian Case, *Politics and Governance*, 14, art. 10692

Regione Emilia-Romagna (2016), *La legislazione regionale in materia di partecipazione. Dati estrapolati dalle schede di rilevazione sugli strumenti di partecipazione deliberativa*, Regione Emilia-Romagna – Servizio Affari legislativi e aiuti di Stato

Rescigno G. U. (2002), Principio di sussidiarietà orizzontale e diritti sociali, *Diritto pubblico*, n.1, pp.5-50

Wilcox D. (2003), *The Guide to Effective Participation*, Brighton, Partnership

Yi C., Qiu X. (2025), From local government to local governance: A systematic literature review and future research agenda, *Australian Journal of Public Administration*, 84, n.1, pp.102-124

Vittoria Laino
vittoria.laino@uniroma1.it
Dottoranda in Studi Politici presso Sapienza Università di Roma, in cotutela con il Dottorato in Democrazia nel XXI secolo dell'Università di Coimbra. Ha lavorato come consulente per il Dipartimento della Funzione pubblica e collaborato con l'*Agência para a Reforma Tecnológica do Estado* (Portogallo). Si occupa di analisi e valutazione delle politiche pubbliche, democrazia partecipativa, istituzionalizzazione della partecipazione ed *empowerment* della cittadinanza. È autrice della monografia *Al di là del tavolo. Processi partecipativi e co-progettazione: modelli e strumenti* (2025).

Scaffale - Rubrica di recensioni

Il coraggio di contare. Storie di donne, finanza ed etica nell'Italia contemporanea

Natascha Lusenti – Il Saggiatore, 2024, pp. 232



Con *Il coraggio di contare*, Natascha Lusenti propone un'opera che attraversa i confini tra saggio, narrazione e inchiesta sociale, ponendo al centro il rapporto tra donne, denaro ed etica. Il volume nasce dall'incontro con il collettivo delle donne di Banca Etica ed esplora, attraverso decine di testimonianze, i nodi che ancora oggi condizionano l'autonomia economica femminile nell'Italia contemporanea. Le storie raccolte mostrano quanto la relazione tra donne e finanza sia tutt'altro che marginale: è, al contrario, uno dei campi in cui più chiaramente si manifestano disuguaglianze strutturali e forme di esclusione sociale. Il percorso tra voci diverse di donne (studentesse, imprenditrici, lavoratrici del terzo settore, psicologhe, socie e dipendenti di Banca Etica) smonta il pregiudizio secondo cui donne e finanza sarebbero mondi lontani e mostra, invece, quanto il denaro possa essere uno strumento di indipendenza, cambiamento sociale e liberazione dalla violenza economica.

Uno dei contributi più rilevanti del libro riguarda infatti la violenza economica, trattata non come un concetto astratto, ma attraverso vissuti concreti: donne escluse dalle decisioni finanziarie familiari, limitate nell'accesso ai beni materiali, talvolta obbligate alla contrazione di debiti o all'abbandono del lavoro. Molte donne intervistate vivono o hanno vissuto situazioni di dipendenza economica, dovuta all'assenza di un conto corrente personale o alla necessità di affidarsi economicamente a familiari o partner. Lusenti rende evidente come tale forma di abuso continui a essere sottovalutata e difficilmente riconosciuta, proprio perché si intreccia a ruoli di genere radicati e mostra le conseguenze sociali e personali della mancanza di autonomia finanziaria.

Il dato narrativo trova consonanza con le recenti analisi e ricerche Inapp[1] sulla dipendenza economica come determinante delle disuguaglianze di genere: la carenza di autonomia finanziaria amplifica vulnerabilità già esistenti e produce effetti sulla partecipazione al lavoro, sulla qualità dell'occupazione e sulla possibilità di accedere a opportunità formative e professionali. La ridotta partecipazione femminile al mercato del lavoro, i divari retributivi e gli stereotipi sociali alimentano una forma strutturale di dipendenza economica che può sfociare, appunto, in violenza economica, spesso difficile da misurare e riconoscere sia a livello statistico che istituzionale.

Lusenti restituisce queste criticità mostrando come i percorsi individuali siano costretti a confrontarsi con un sistema che attribuisce ancora un valore economico insufficiente al lavoro femminile, soprattutto a quello di cura. Le testimonianze sono restituite quasi in forma poetica, con una struttura frammentata che lascia spazio al non detto e alle emozioni; le voci raccolte dall'Autrice non sono soltanto esempi ma personaggi narrativi a tutti gli effetti che raccontano frammenti di vita e parlano di maternità e lavoro, precarietà, aspirazioni tradite, scelte difficili, desiderio di riscatto, paura di non valere, necessità di essere accettate, mettendo al centro l'indipendenza economica come prerequisito della libertà femminile.

L'Autrice ricorda che il denaro è un fattore culturale e sociale, non solo materiale e propone una visione della finanza come bene comune e non come spazio esclusivo; invita a interrogarsi su cosa significhi "contare", in tutti i sensi del termine: contare i soldi, contare come persone, contare per sé stesse. È un libro che parla soprattutto alle donne ma che mette in discussione strutture e abitudini collettive, rappresentando un invito a riconoscere il valore del proprio lavoro, della propria voce e della propria indipendenza economica. Le

1 Chiurco L. (2025), *La dipendenza finanziaria in prospettiva di genere. Riconoscere e contrastare la violenza economica*, Inapp Paper n.59, Roma, Inapp.
Inapp, Calia R.M., Di Iorio T., Ferrara M., Parente M., Porcarelli C. (2026), *Valorizzare gli interventi di empowerment a favore delle donne vulnerabili. Primi esiti e riflessioni*, Inapp Report n. 68, Roma, Inapp.

esperienze raccolte mostrano come la finanza possa diventare uno strumento di coesione e cambiamento sociale quando orientata a modelli cooperativi, inclusivi e finalizzati al bene comune. Questo approccio risuona con alcune linee di ricerca Inapp dedicate alle politiche del lavoro e dell'inclusione sociale, che valorizzano il ruolo della comunità e dell'empowerment nell'affrontare vulnerabilità economiche e sociali.
La struttura del libro - ibrida, intima, a tratti poetica - permette di intrecciare le testimonianze contemporanee con riferimenti a pensatrici come Judith Butler, Audre Lorde, Anna Bravo e alla Nobel Claudia Goldin[2], creando una genealogia del pensiero femminista attorno al valore del lavoro e del denaro. Tale scelta narrativa amplifica la densità del testo e ne rafforza la connessione con le più recenti riflessioni sociologiche e politiche sulla giustizia economica e di genere.
Il coraggio di contare offre un contributo significativo al dibattito pubblico sull'autonomia economica delle donne, mettendo in luce dinamiche che la ricerca empirica, compresa quella dell'Inapp, conferma come centrali nel quadro delle disuguaglianze di genere. Il testo costituisce un prezioso riferimento sia per la capacità di restituire la realtà attraverso il vissuto, sia per l'allineamento con le analisi istituzionali sulle condizioni di vulnerabilità economica femminile.

Cristiana Porcarelli
INAPP

2 Con riferimento all'opera di queste autrici, si veda, in particolare: Bravo A. (2005), *A colpi di cuore,* Torino, Bollati Boringhieri; Butler J. (1990), *Gender Trouble*, New York, Routledge; Goldin C. (2021), *Career and Family*, Princeton, Princeton University Press; Lorde A. (1984), *Sister Outsider*, Freedom, (CA), Crossing Press.

Diritto o privilegio? Fabbisogni abitativi e politiche locali per la casa

Annalisa Gramigna e Antonello Scialdone (*a cura di*) – Egea, 2026, pp. 304



La domanda posta dal titolo del volume curato da Gramigna e Scialdone è solo in apparenza retorica. La casa è riconosciuta come condizione essenziale della cittadinanza, ma per una parte crescente della popolazione italiana è un bene selettivo, accessibile in funzione del reddito, del patrimonio familiare e della posizione occupazionale. Il libro assume questa contraddizione come punto di partenza e ricompone temi spesso trattati separatamente: edilizia pubblica, aumento dei costi, finanziarizzazione immobiliare, trasformazioni familiari, studenti, invecchiamento e ruolo dei Comuni.

Il volume nasce dalla collaborazione tra Ifel e Inapp ed è un'opera collettiva e multidisciplinare. Sociologia, economia, urbanistica e analisi delle politiche pubbliche concorrono a una lettura dell'abitare che supera l'identificazione tra politica della casa e produzione edilizia. La tesi comune è che non basti aumentare il numero degli alloggi: occorre modificare il modo in cui il patrimonio viene prodotto, distribuito, utilizzato e governato, mettendo in relazione casa, reddito, servizi, lavoro, salute e qualità urbana.

La prima parte ricostruisce l'evoluzione del bisogno e i problemi di accessibilità. Il saggio iniziale di Francesca Gelli fornisce la cornice politica dell'intero volume. La lunga marginalità della questione abitativa nell'agenda nazionale non è interpretata come una semplice dimenticanza, ma come l'esito di inerzie, interessi consolidati e processi di 'non decisione'. Il decentramento avviato negli anni Novanta, pur favorendo alcune sperimentazioni, ha coinciso con il ritiro dello Stato, la riduzione delle risorse e la perdita di competenze tecniche. Ne è derivato un sistema policentrico ma privo di una strategia nazionale, nel quale la capacità di risposta dipende eccessivamente dal territorio.

Particolarmente rilevante è il contributo di Marianna Filandri e Violetta Tucci sulle disuguaglianze e sulla finanziarizzazione. Le autrici mostrano come il modello italiano, fondato sulla diffusione della proprietà, sia stato costruito da politiche fiscali, creditizie e urbanistiche che hanno favorito la casa come patrimonio prima che come servizio sociale. La proprietà ha protetto molte famiglie, ma ha anche trasformato le differenze ereditarie in disuguaglianze abitative. Chi dispone di risparmi, immobili o garanzie familiari può acquistare e trasmettere ricchezza; chi ne è privo resta confinato in un mercato dell'affitto più instabile. La finanziarizzazione accentua la frattura, trattando l'alloggio come un asset e subordinandone la funzione d'uso al rendimento.

Da qui una delle indicazioni più convincenti del libro: spostare l'attenzione dall'oggetto-casa ai soggetti che la abitano. L'accessibilità non è una qualità intrinseca dell'immobile, ma una relazione tra costo, reddito, composizione familiare e continuità nel tempo. Anche il capitolo di Francesco Belletti e Sara Nanetti amplia lo sguardo, definendo la casa una 'infrastruttura relazionale', il luogo in cui si organizzano cura, reciprocità e rapporti tra generazioni. Il sostegno familiare consente a molti italiani di accedere all'alloggio, ma supplisce alla debolezza delle politiche pubbliche e riproduce la distanza tra chi può contare su eredità e chi ne è escluso. La qualità dell'abitare dipende inoltre dal quartiere, dai servizi e dalle reti di prossimità.

I capitoli quantitativi restituiscono bene la specificità della crisi italiana. I valori medi descrivono un Paese apparentemente protetto dall'alta quota di proprietari e da costi non sempre superiori a quelli europei. Ma la media nasconde sovraffollamento, povertà energetica, morosità e rischio di sfratto. Si forma così una vasta 'fascia grigia': famiglie troppo ricche per accedere all'edilizia pubblica, ma troppo povere o precarie per sostenere il mercato. A questa domanda mobile corrisponde un patrimonio pubblico ridotto e spesso vetusto, oltre a criteri di accesso che non rappresentano più la pluralità dei bisogni. Giovani precari, nuclei monogenitoriali, anziani soli, migranti e lavoratori mobili restano frequentemente fuori dalle griglie tradizionali.

Uno dei meriti del volume è mostrare che la crisi non dipende soltanto dalla scarsità di risorse, ma anche dalla debolezza dell'infrastruttura conoscitiva e amministrativa. I dati sono frammentati e i Comuni conoscono spesso il bisogno solo attraverso le domande rivolte ai servizi esistenti. Si misura ciò per cui è già prevista una risposta, mentre resta invisibile ciò che non entra nelle categorie amministrative. Senza una conoscenza dinamica dello stock, degli usi, dei redditi e dei costi, la politica insegue l'emergenza e fatica a valutare gli interventi.

La seconda parte si concentra sulle politiche locali e sulle sperimentazioni. Il capitolo di Gramigna restituisce la condizione dei Comuni, chiamati a rispondere all'emergenza, alla gestione dell'ERP e ai bisogni della fascia grigia. La riduzione degli investimenti ha spostato l'azione pubblica dalla costruzione alla gestione sociale del disagio, caricando i servizi locali di funzioni riparative. Il welfare abitativo è indispensabile, ma non sostituisce una politica strutturale; anche un nuovo programma di edilizia pubblica sarebbe insufficiente se non coordinato con lavoro, regolazione delle locazioni, fiscalità, servizi e pianificazione urbana.
I casi di Foggia, Parma e Torino mostrano la vitalità dell'innovazione locale. Il progetto di Foggia lega il diritto allo studio al diritto alla città. Il programma *Fa' la casa giusta!* di Parma prova a mettere in sistema urbanistica, welfare e patrimonio pubblico e privato. *Homes4All*, nato a Torino, sperimenta invece l'acquisizione e il recupero di immobili sottoutilizzati attraverso partenariati e finanza a impatto sociale. In questi contesti, la casa diventa il punto di accesso a una presa in carico più ampia e il gestore sociale un mediatore tra abitanti, proprietari, servizi e comunità.
Il contributo più radicalmente propositivo è quello di Luigi Corvo e Andrea Battistoni, quello dedicato all'housing capitalizzabile. L'idea è trasformare una parte del canone pagato dall'inquilino in capitale personale, trasferibile o utilizzabile per l'acquisto dell'alloggio. Si tenta così una terza via tra affitto, inteso come spesa priva di accumulazione, e mutuo, sempre meno accessibile a chi ha redditi discontinui e non dispone di garanzie familiari. La proposta affronta un nodo rimosso: in una società patrimoniale, la disuguaglianza non riguarda soltanto quanto si paga per abitare, ma la possibilità di trasformare la spesa in sicurezza futura. Restano da verificare la sostenibilità nei mercati più costosi, la distribuzione dei rischi e le garanzie affinché l'obiettivo sociale non venga subordinato al rendimento.
Qui emerge una tensione che avrebbe meritato una discussione più sistematica. Da un lato, molti saggi denunciano la trasformazione della casa in asset finanziario e i processi estrattivi che espellono popolazione dalle aree attrattive. Dall'altro, diverse soluzioni fanno affidamento su fondi, investitori privati e finanza d'impatto. La contraddizione non rende le proposte incoerenti, ma impone di precisare quali regole distinguano una finanza 'civile' dalla finanziarizzazione: limiti ai rendimenti, durata dei vincoli sociali, controllo pubblico, distribuzione del rischio, trasparenza e partecipazione degli abitanti. Senza tali condizioni, il partenariato rischia di ridurre il rischio privato con risorse pubbliche senza produrre un patrimonio permanentemente accessibile.
I dati raccolti nel volume fanno emergere una contraddizione decisiva del modello italiano: l'ampiezza del patrimonio abitativo e la diffusione della proprietà non garantiscono, di per sé, un accesso equo alla casa. Il problema non consiste soltanto nella quantità degli alloggi disponibili, ma nella loro distribuzione, nelle forme d'uso e nella crescente distanza tra i costi dell'abitare e le risorse delle famiglie. Accanto a nuclei protetti dalla proprietà e dai trasferimenti intergenerazionali, cresce una popolazione che dipende interamente dal reddito corrente e che, pur non rientrando nelle condizioni di povertà estrema, non riesce ad accedere stabilmente né alla proprietà né alla locazione di mercato. Il volume mostra così come le categorie tradizionali delle politiche pubbliche intercettino soltanto una parte del disagio, lasciando senza risposta bisogni sempre più differenziati e mutevoli. La questione centrale diventa allora non solo ampliare l'offerta, ma ripensare destinatari, criteri di accesso e strumenti di intervento alla luce delle trasformazioni demografiche, lavorative e familiari.
Su questo punto il libro approda nelle pagine finali. La questione abitativa manifesta un doppio disallineamento: tra la casa e l'organizzazione del lavoro e dei servizi; tra la dimensione del bisogno e la scala delle risorse. Per superarlo occorre una regia nazionale che non annulli le autonomie locali, ma stabilisca obiettivi, standard, strumenti finanziari e infrastrutture informative comuni. L'abitare non è una politica settoriale: è una componente strutturale del welfare e, insieme, una politica urbana. Decide chi può restare nelle città, studiare, lavorare, formare una famiglia, invecchiare in autonomia e partecipare alla vita collettiva.
Diritto o privilegio? non offre una ricetta unica, né trasforma sempre la ricchezza delle proposte in un'agenda operativa coerente. Offre però qualcosa che manca nel dibattito italiano: una rappresentazione multidimensionale della crisi e un lessico comune tra studiosi, amministratori e operatori. Il suo contributo principale consiste nel sottrarre la casa tanto alla sola retorica emergenziale quanto alla neutralità del mercato. La casa è un'infrastruttura di cittadinanza, capace di riprodurre privilegi oppure di redistribuire opportunità. Finché l'accesso a un'abitazione adeguata dipenderà dal patrimonio ereditato, dalla famiglia di origine o dalla capacità di offrire garanzie al mercato, la casa resterà formalmente un diritto e sostanzialmente un privilegio.

Enrico Puccini
Osservatorio Casa Roma

Per una sociologia del confine. Spazio e identità nella società contemporanea

Ilaria Riccioni – Carocci, Roma, 2025, pp. 104



Il volume *Una sociologia del confine* di Ilaria Riccioni propone una riflessione articolata e convincente su uno dei concetti più rilevanti per comprendere le trasformazioni della società contemporanea: il confine. La riflessione dell'Autrice, che insegna Sociologia a Bolzano, si colloca nel più ampio dibattito delle scienze sociali, che ha progressivamente superato una visione esclusivamente geografica dei confini. L'obiettivo di questa studiosa non è quello di offrire una trattazione esaustiva o manualistica, bensì di proporre, all'interno del quadro sociologico, un'analisi assai ricca che interpreta il confine non come un elemento statico ma come un dispositivo sociale, simbolico e politico, superando la sua tradizionale riduzione a mera linea territoriale. L'Autrice, infatti, esplora le molteplici declinazioni del concetto di confine e di limite, nelle loro dimensioni teoriche e operative, mostrando come esse attraversino tanto l'esperienza quotidiana quanto la ricerca scientifica.

Nel testo il confine non è inteso esclusivamente come soglia o linea di separazione, ma come concetto utile a individuare contesti distinti e relativamente autonomi, capaci di entrare in relazione con gli spazi circostanti a livello globale e sistemico. Se in origine lo studio dei confini era prevalentemente legato allo spazio territoriale, oggi l'attenzione si estende ai fenomeni di mobilità e migrazione, evidenziando il ruolo dei confini nella regolazione dei flussi e nella costruzione delle appartenenze sociali. Il confine viene inteso come prodotto sociale, risultato di processi storici e relazionali, capace di generare inclusioni ed esclusioni, appartenenze e alterità.

In questa prospettiva, Riccioni analizza l'idea di un mondo 'senza confini' prodotta dalla globalizzazione, mostrando come, al contrario, i confini persistano anche oggi e si riorganizzino assumendo nuove forme e funzioni. Le trasformazioni politiche degli ultimi decenni, infatti, non riguardano soltanto i modelli istituzionali, ma anche i processi attraverso cui il potere viene esercitato e legittimato. Particolarmente significativo è il riferimento al temporaneo ripristino delle frontiere interne europee durante la pandemia, che conferma il ruolo del confine come dispositivo di controllo e gestione dell'alterità.

Il volume è arricchito da esempi empirici di confini, come il caso dell'Alto Adige/Sud Tirolo, territorio di frontiera e provincia autonoma, o del Mar Mediterraneo, spazio che oscilla tra funzione di culla di scambi culturali e ruolo di frontiera politica. Questi esempi mostrano come i confini siano da considerarsi quali spazi complessi e interconnessi con le pratiche sociali, piuttosto che semplici linee fisiche o amministrative. Il confine emerge così come costruzione sociale e culturale che incide profondamente sui processi di identità, appartenenza ed esclusione.

Uno degli elementi centrali dell'argomentazione riguarda la dimensione ambivalente del confine, che opera simultaneamente come dispositivo di separazione e come spazio di contatto, negoziazione e trasformazione: i confini possono agire sia come barriere sia come "*ponti e punti di incontro tra culture diverse*" (pag.14). In questo senso, il confine non è soltanto ciò che divide, ma anche ciò che rende possibile l'incontro e la produzione di identità, ridefinendo continuamente appartenenze e alterità. Come sottolinea l'Autrice: "*l'essere umano è definito dai confini e da come questi confini vengono modellati, in quanto diventano limiti dell'essere, dell'agire e del pensare*" (pag.21).

Nel volume un'attenzione particolare viene riservata, inoltre, all'analisi del rapporto tra confine e identità. L'Autrice sottolinea come il confine diventi uno strumento attraverso cui le società definiscono chi è incluso e chi è escluso, chi appartiene a un 'noi' e chi viene percepito come 'altro', incidendo sui processi di costruzione dell'appartenenza sociale, culturale e politica. In questo senso, il confine assume una forte valenza simbolica, che si manifesta non solo a livello geopolitico, ma anche nella vita quotidiana, nei confini sociali, culturali e relazionali.

Il volume dedica inoltre uno spazio significativo al contesto europeo, mostrando come l'Unione europea rappresenti un laboratorio emblematico di ridefinizione dei confini: alcuni vengono attenuati o resi invisibili,

mentre altri vengono rafforzati, soprattutto in relazione alle migrazioni e alla mobilità. Nel complesso, *Una sociologia del confine* si presenta come un saggio ricco di contenuti che invita a guardare in modo critico e consapevole alle molteplici barriere — visibili e invisibili — che attraversano la società contemporanea. Il libro si configura così come uno spazio critico di riflessione, utile per comprendere come i confini continuino a ridefinire l'agire sociale e le forme di appartenenza nella società contemporanea.
Nell'offrire strumenti concettuali utili per interpretare la complessità del presente, il contributo di Riccioni risulta particolarmente pregnante in un contesto storico quale quello attuale, segnato da migrazioni, conflitti identitari e ridefinizioni politiche. La natura innovativa del volume, che emerge da un'esplorazione trasversale dei concetti di limite e di confine, nei diversi profili disciplinari, geografici, identitari e simbolici, si accompagna a una riflessione sulla dimensione metodologica, che assume una connotazione marcatamente qualitativa. Attraverso questa prospettiva, il libro invita a ripensare strumenti e categorie di analisi dei fenomeni sociali, aprendo a letture inedite delle trasformazioni in atto.

Simona Testana
INAPP

Per proporre un articolo

La Rivista pubblica articoli sui temi legati a monitoraggio, analisi e valutazione delle politiche del lavoro, dell'istruzione, della formazione, delle politiche sociali e, in generale, tutte le politiche economiche che hanno effetti sul mercato del lavoro.
Sinappsi pubblica solo articoli inediti. I contributi non possono pertanto riguardare articoli già pubblicati, anche solo in parte, su altre riviste italiane e straniere o essere contemporaneamente proposti ad altre riviste per la pubblicazione. I contributi possono essere proposti in lingua italiana o in lingua inglese. Gli articoli devono essere inviati in formato word all'indirizzo di posta elettronica sinappsi@inapp.gov.it.
I testi vanno corredati con la scheda di accompagnamento contenente la dichiarazione, sotto propria responsabilità, di originalità della proposta (https://www.inapp.gov.it/rivista-sinappsi/indicazioni-per-gli-autori) e dagli allegati (vedi infra).

Procedure

Ogni proposta, dopo la verifica della presenza dei requisiti minimi di pubblicabilità (rispetto delle norme editoriali) è soggetta all'esame preliminare del Comitato editoriale. Se giudicato coerente con gli obiettivi e gli standard qualitativi della Rivista, il testo è sottoposto, in forma anonima, al giudizio di almeno un referee in *double blind peer review* (ovvero autori e revisori reciprocamente anonimi). Il testo inviato ai revisori non deve pertanto contenere informazioni sull'identità degli autori. Questi sono quindi tenuti a minimizzare le autocitazioni e qualsiasi altro connotato che possa favorire la loro identificazione da parte dei revisori.

Il processo di revisione da parte dei referee potrà dar luogo a uno dei seguenti esiti: accettazione; accettazione subordinata a modifiche minori; accettazione subordinata a modifiche rilevanti; da sottomettere a riesame previe modifiche e senza impegno di successiva accettazione; rifiuto. L'accettazione subordinata a modifiche prevede la revisione da parte degli autori, che dovranno rendere evidenti nel testo le modifiche effettuate. Quando l'articolo è accettato per la pubblicazione, gli autori trasferiscono automaticamente all'Inapp ogni diritto di copyright, garantendo la possibilità della più ampia diffusione.

Agli autori sarà consegnata la prima bozza per la correzione, con l'invito a restituirla entro una data prefissata. Sulla prima bozza potranno essere apportate solo modifiche marginali. La correzione della seconda bozza sarà eseguita a cura della Redazione.

Utilizzo dell'Intelligenza artificiale

L'uso dell'Intelligenza artificiale (IA) deve essere dichiarato in modo trasparente, rispettando i diritti fondamentali di veridicità e tracciabilità, assicurando che il testo sia sviluppato in modo da salvaguardare la privacy dei dati, la proprietà intellettuale, garantendo accuratezza e sostenibilità e controllando a questo scopo i termini e le condizioni di qualsiasi strumento di intelligenza artificiale utilizzato. Gli autori si assumono la piena responsabilità dell'integrità dei dati e dell'autenticità di tutti i contenuti generati o modificati tramite IA.

Dimensione e criteri di stesura dei testi:

- pagina formato A4;
- da 30.000 a 50.000 caratteri complessivi per articolo, spazi inclusi, comprese tabelle e figure (testi di dimensione superiore devono essere concordati con la Redazione, indicando i motivi per cui non è possibile rispettare i limiti previsti);
- titolo max 50 caratteri, eventuale sottotitolo max 70 caratteri, ma se è presente il sottotitolo, il titolo può avere un massimo di 30 caratteri;

- paragrafi numerati (solo primo livello); la Redazione può intervenire sul titolo proponendo modifiche agli autori;
- numero di tabelle + figure non superiore a 10 (non duplicare le informazioni fornite dalle tabelle con quelle dei grafici e del testo dell'articolo);
- tabelle e figure sempre numerate (ad esempio Tabella 1, Tabella 2 ecc.; Figura 1, Figura 2 ecc.), con titolo, fonte e anno. Possono essere utilizzati colori;
- note esplicative inserite a piè di pagina;
- richiami bibliografici inseriti nel testo entro parentesi tonde, con l'indicazione del cognome dell'autore da citare, seguito dalla data della pubblicazione originale ed eventualmente dalla/e pagina/e di riferimento della citazione riportata nel testo (Cognome autore data, numero pagina). A ogni richiamo deve corrispondere la fonte completa in bibliografia, inserita a fine saggio (vedi Norme bibliografiche);
- lo stesso sistema di rinvio alla bibliografia finale si adotta all'interno delle note a pie' di pagina.

Allegati

L'articolo va corredato da:

- una breve nota biografica (circa 600 caratteri spazi inclusi), elaborata in base al seguente modello: "Ricercatore/ trice, assegnista (oppure insegna) presso l'Istituto/Università (Denominazione). Aggiungere eventualmente, altri incarichi di prestigio. Fra le pubblicazioni recenti si segnalano: (indicare un max di due lavori). Indirizzo e-mail";
- dichiarazione, sotto propria responsabilità, di originalità della proposta (presente all'interno della scheda di accompagnamento);
- abstract in italiano e abstract in inglese di max 600 caratteri ciascuno, spazi inclusi;
- tre parole chiave in italiano e tre corrispondenti keyword in inglese;
- il file in formato Excel delle figure e dei grafici inseriti anche nel testo, un elemento per foglio, con numerazione corrispondente a quanto indicato nell'articolo.

Norme bibliografiche

Requisiti della bibliografia

- deve essere unica e collocata alla fine del lavoro;
- deve indicare esclusivamente le opere citate nel testo e nelle note ed essere aggiornata;
- deve prevedere l'ordine alfabetico per cognome dell'autore o del curatore, del primo autore o curatore nel caso di più nomi, e l'ordine cronologico di pubblicazione delle opere dalla più recente alla meno recente (per opere dello stesso autore pubblicate nello stesso anno, si usino le indicazioni a, b, c);
- i lavori di più autori vanno riportati con tutti i nomi.

Monografie

Autori:

Cognome autore e iniziali puntate del nome (anno tra parentesi), *Titolo del volume in corsivo. Se è presente, il sottotitolo va sempre in corsivo preceduto dal punto*, Luogo, Editore

Nel caso di più autori, mettere tutti gli autori separati da virgole.

Curatori:

Cognome curatore e iniziali puntate del nome (a cura di) (anno tra parentesi), *Titolo del volume in corsivo. Se è presente, il sottotitolo va sempre in corsivo preceduto dal punto*, Luogo, Editore

Per i testi stranieri mettere (eds.) al posto di (a cura di) nel caso di più curatori, (ed.) nel caso di curatore unico.

Nel caso di più curatori, mettere tutti i curatori separati da virgole.

Se la monografia fa parte di una collana, inserire nome della collana e relativo numero dopo il titolo.

Articoli di riviste/periodici

Cognome autore e iniziali puntate del nome (anno tra parentesi), Titolo dell'articolo in tondo. Se è presente, il sottotitolo va preceduto dal punto, Titolo del periodico/rivista in corsivo, annata[1], numero anno reso con n. e numero in cifre, pagine di inizio e fine articolo reso con pp. …-… (senza spazio dopo il punto. Es.: pp.33-45).

Nel caso di più autori, mettere tutti gli autori. Se presente, inserire il DOI tra parentesi uncinate < > senza spazi dopo e prima.

Estratti da monografie

Cognome autore e iniziali puntate del nome (anno tra parentesi), Titolo dell'estratto in tondo. Se è presente, il sottotitolo va preceduto dal punto, in Cognome autore e iniziali puntate del nome, *Titolo del volume in corsivo*, Luogo, Editore, pagine di inizio e fine articolo reso con pp. …-… (senza spazio dopo il punto. Es.: pp.40-60).

Nel caso di più autori, mettere tutti gli autori.

Se il volume di estrazione è a cura di, seguire le indicazioni per i volumi con curatore.

Testi Inapp

I testi Inapp seguono le indicazioni precedenti. Le monografie però devono SEMPRE riportare Inapp fra gli autori o curatori.

1 Annata: insieme di fascicoli di un periodico pubblicati nel corso di un anno o di un periodo editoriale determinato. Fonte http://elearning.unimib.it/mod/glossary/view.php?id=13076.

TESTI INAPP

Monografie

Inapp, Checcucci P., Fefè R., Scarpetti G. (a cura di) (2017), *Età e invecchiamento della forza lavoro nelle piccole e medie imprese italiane*, Inapp Report n.3, Roma, Inapp

Paper

Quaranta R., Ricci A. (2017), *Riforma delle pensioni e politiche di assunzione*. *Nuove evidenze empiriche, italiane*, Inapp Paper n.3, Roma, Inapp

Sinappsi

Cassese S. (2018), Evoluzione della normativa sulla trasparenza, *Sinappsi*, VIII, n.1, pp.5-7

Letteratura grigia

Schulz M., Bressers D., van der Steen M., van Twist M. (2015), Internal Advisory Systems in Different Political-Administrative Regimes, *Prepared for the International Conference on Public Policy (ICPP) T08P06 – Comparing policy advisory systems at the second International Conference on Public Policy, Milan 2015*

Comité de suivi du Cice, France Stratégie (2016), Comité de suivi du Crédit d'impôt pour la compétitivité et l'emploi. Rapport 2016, *Evaluation, Septembre* 2016

Giurisprudenza

Corte costituzionale

Corte cost. 25 luglio 1995 n.376, in *Giur. cost*., 1995, XL, 4, p.2750 ss.

Corte di Cassazione

Cass., sez. III, 14 ottobre 1991 n.10763, in *Dir. Trasp*., 1993, VI, 3, p.847 ss.

Cass. pen., sez. un., 26 marzo 2003, in *Cass. pen*., 2003, XLIII, 9, p.2579 ss.

Cass. pen., sez. VI., 3 novembre 2001, in *Riv. pen*. 2002, 1, p.31 ss.

Cass. civ., sez. lavoro, 29 maggio 1998, n.5348 Cass. pen., sez. I, 30 aprile 1992, Idda, in *C.E.D. Cass.* Pen., n.190564

Cass. pen., sez. un., 6 novembre 1992, Martin, in *Cass. Pen*., 1993, XXXIII, 2, p.280

Cass. pen., sez. IV, 21 ottobre 2005, in *Dir. Pen. Proc*., 2006, XII, 2, p.200

Cass. pen., sez. V, d 24 ottobre 2002, De Vecchis, in *Guida dir.*, 2003, X, 10, p.86

Consiglio di Stato

Cons. Stato, sez. IV, 14 giugno 2005 n.3120, in *Foro Amm. CDS*, 2005, IV, 6, p.1728 ss.

Corte dei conti

Corte conti 16 luglio 2010 n.15, in *Riv. corte conti,* 2012, LXV, 3-4, p.10

Corte d'Appello

App. Napoli 3 novembre 2008, in *Foro it*., 2009, CXXXIV, 5, pt. I, p.1476 ss.

Corte d'Assise

Corte Assise Milano 15 febbraio 2006, in *Giur. merito*, 2007, XXXIX, 3, p.783 ss., con nota di L.D. CERQUA

Tribunale

Trib. Roma 27 giugno 2005, in *Lavoro nella giur*., 2007, XV, 3, p.283, con nota di B. DE MOZZI

Tribunale Amministrativo Regionale

Tar Bari Puglia 6 aprile 2005 n.1376, in *Foro amm*.TAR, 2005, IV, 4, p.1214

Pretura

Pretore di Gubbio ord. 12 febbraio 1957, in Giur. cost., 1957, II, 1, p.127 ss.

Corte di Giustizia dell'Unione europea

Corte Giust., 28 giugno 1978, C-70/77, Simmenthal c. Amministrazione delle Finanze, in Racc., 1978, p.453

Letteratura grigia

La letteratura grigia segue le precedenti indicazioni rispetto al metodo Autore/Data. È necessario riportare sempre tutti gli elementi utili a rintracciare la pubblicazione:
Autori/Ente autore (anno), titolo del contributo, *informazioni aggiuntive*. Se disponibili, riportare il link al documento e/o il DOI tra parentesi uncinate < > senza spazi dopo e prima.

Giurisprudenza

Organo giurisdizionale emanante (Cassazione, Tribunale, Consiglio di Stato), tipo di atto adottato (Sentenza, Ordinanza, Decreto), sezione dell'organo emanante (non sempre presente), data della pronuncia, numero o nome delle parti (non sempre previsto. Se c'è il nome della parte dopo la data si tratta di un provvedimento della giurisdizione penale).

Legislazione

In ogni capitolo la prima citazione deve essere completa.
(Es. D.P.R. 26 luglio 1976 n.752, Norme di attuazione dello statuto speciale della Regione Trentino-Alto Adige in materia di ...)
Le citazioni successive possono essere in forma abbreviata.
(Es. D.P.R. n.752/1976)
La citazione degli articoli deve consentire l'individuazione precisa della disposizione normativa.
(Es. art. 5, comma 2, D.P.R. n.752/1976)
Citazioni all'interno del testo: legge n.150/2000, oppure L. n.150/2000

Risorse elettroniche

Le risorse elettroniche seguono le indicazioni precedenti rispetto al metodo Autore/Data. È SEMPRE necessario mettere il link al testo e/o pagina web di riferimento. Per le pagine inserire la dicitura (consultato il ….).

ESEMPI

Monografie

Campbell J.L., Pedersen O.K. (2014), *The national origins of policy ideas. Knowledge regimes in the United States, France, Germany and Denmark*, Princeton, Princeton University Press

Facchini C. (a cura di) (2008), *Conti aperti. Denaro, asimmetrie di coppie e solidarietà tra le generazioni,* Bologna, Il Mulino

Eichbaum C., Shaw R. (eds.) (2010), *Partisan Appointees and Public Servants, an International Analysis of the Role of the Political Adviser,* Cheltenham UK, Edward Elgar Publishing Limited

Eichhorst W., Wintermann O. (2005), *Generating Legitimacy for Labor Market and Welfare State Reforms. The Role of Policy Advice in Germany, the Netherlands and Sweden*, IZA Discussion Paper n.1845, Bonn, IZA

Articoli di riviste/periodici

Craft J., Halligan J. (2017), Assessing 30 years of Westminster policy advisory system experience, *Policy Sciences*, 50, n.1, pp.47-62 <DOI 10.1007/s11077-016-9256-y>

Estratti da monografie

Pattyn V., van Voorst S., Mastenbroek E., Dunlop C. A. (2017), Policy evaluation in Europe, in Ongaro E., Van Thiel S., *The Palgrave Handbook of Public Administration and Public Management*, Bristol, Policy Press, pp.105-11

Corte internazionale di Giustizia

Corte internazionale di Giustizia, sentenza del 27 giugno 1986, Attività militari e paramilitari contro il Nicaragua

Corte penale internazionale

Corte penale internazionale, Prima Camera di I grado, 14 marzo 2012, Thomas Lubanga Dyilo

Corte europea dei Diritti dell'Uomo

C. eur. Dir. Uomo, 12 febbraio 2013 – Ricorso n.24 818/03 – causa Armando Iannelli c. Italia

C. eur. Dir. Uomo, sentenza del 24 ottobre 1986, nel caso Agosi contro Regno Unito

Legislazione

- D.L. 27 giugno 1997 n.185
- D.M. 5 marzo 1999
- D.Lgs. 29 marzo 1993 n.119
- L. 13 febbraio 2001 n.45
- Art. 456 c. c.
- Art. 16, comma 4, lett. a, L. 28 gennaio 1994 n.84
- Art. 1 reg. CEE n.4056/86 del 22 dicembre 1986
- Regolamento n.1254/2008/CE della Commissione, che modifica il regolamento (CE) n.889/2008 recante modalità di applicazione del regolamento (CE) n.834/2007 del Consiglio relativo alla produzione biologica e all'etichettatura dei prodotti biologici, per quanto riguarda la produzione biologica, l'etichettatura e i controlli, in GU L 337 del 16.12.2008
- Direttiva n.70/50/CEE della Commissione, 22 dicembre 1969, in GUCE L 13, 19.1.1970

SINAPPSI è la rivista scientifica dell'Inapp (Istituto nazionale per l'analisi delle politiche pubbliche), ente di ricerca che svolge analisi, monitoraggio e valutazione delle politiche, in particolare di quelle che hanno effetti sul mercato del lavoro. Occupazione, istruzione, formazione e welfare sono i principali temi di studio. La rivista intende intrecciare, connettere appunto, i risultati e i contenuti della ricerca, per offrire al lettore una rete di riflessioni e ai decisori politici un impulso per le scelte strategiche.

ISSN 2532-8549

INAPP

ISBN: 978-88-543-0390-4