

Convegno Inps-Siep  
I dati amministrativi per le analisi socio-economiche e la valutazione delle  
politiche pubbliche, *Roma, 23 maggio 2018*

Sessione I  
L'accesso e la disponibilità dei dati amministrativi

# Accesso e utilizzo di dati amministrativi: strategia, risultati e prospettive della statistica ufficiale

---

Roberto Monducci

ISTITUTO NAZIONALE DI STATISTICA

## Contenuto della presentazione

- L'utilizzo di dati amministrativi nella statistica ufficiale
- L'esperienza dell'Istat
- I vantaggi del modello integrato: un esempio
- Prospettive: estensione del modello basato sui registri

## L'utilizzo di dati amministrativi nella statistica ufficiale (1)

Progressiva affermazione di un modello di produzione di statistiche ufficiali basato sull'utilizzo integrato di fonti statistiche e amministrative

- ▶ **Approccio europeo**: modello incorporato formalmente nei regolamenti statistici europei (art 24. Reg. CE N. 223/2009 del Parlamento e del Consiglio), approvati dagli organi rappresentativi attraverso procedure complesse e rigorose, e verificato dagli organismi di controllo.
- ▶ **Vantaggi**:
  - semplificazione e **riduzione degli adempimenti** di cittadini e imprese;
  - **riduzione dei costi** per la Pubblica amministrazione per la funzione statistica (es. Censimento permanente della popolazione ...);
  - **aumento dell'accuratezza** delle stime (uso di basi dati esaustive).
  - **aumento della capacità informativa** dell'output.

## L'utilizzo di dati amministrativi nella statistica ufficiale (2)

- ▶ **Revisione** del tradizionale processo di produzione statistica e costruzione di nuovi strumenti metodologici e di valutazione della qualità dei processi e dei dati (**Eurostat, 2003**).
- ▶ **Investimenti in metodologie avanzate** e **infrastrutture** per la trasformazione dei dati amministrativi in informazione statistica (classificazioni, coverage, variabili, imputazione, stima ...).
- ▶ Creazione di **reti di ricerca metodologica** a livello europeo per identificare soluzioni avanzate per il trattamento dei dati amministrativi.
- ▶ Risultati significativi raggiunti dalla statistica europea nel campo sia delle **statistiche economiche** (imprese, domanda di lavoro e retribuzioni, commercio estero ecc.) sia delle **statistiche demografiche e sociali** (demografia, salute, istruzione ecc.).

## L'utilizzo di dati amministrativi nella statistica ufficiale (3)

### Programma Statistico Nazionale: crescita tendenziale dell'utilizzo di dati amministrativi a fini statistici

	Statistiche da indagine (SDI)	Statistiche da fonte amministrativa (SDA)
2013	376	206
2016	327	222
2019	310	275

- ▶ Tra il 2013 e il 2019 la raccolta diretta di dati da famiglie, unità economiche e istituzioni diminuisce da 376 a 310 lavori (-17,6%);
- ▶ Contestualmente aumentano da 206 a 275 gli utilizzi di fonti amministrative (+33,5%), spesso in combinazione con fonti statistiche dirette.

## L'utilizzo di dati amministrativi nella statistica ufficiale (4)

### Enti, archivi, utilizzi: la pianificazione PSN 2019

- ▶ Nel 2019 275 SDI e 310 SDA utilizzeranno 1.043 volte 478 archivi amministrativi di 94 enti fornitori.
- ▶ Gli **archivi** maggiormente utilizzati:
  - Anagrafi comunali; Uniemens; Anagrafe degli studenti; Anagrafe e modelli tributari
- ▶ Dopo l'Istat (poco meno del 60% degli utilizzi) i maggiori **utilizzatori** di archivi di altre fonti sono:
  - Ministero del lavoro; Inps; Istituto superiore di sanità; Ministero dell'Istruzione; Comuni capoluogo di regione.

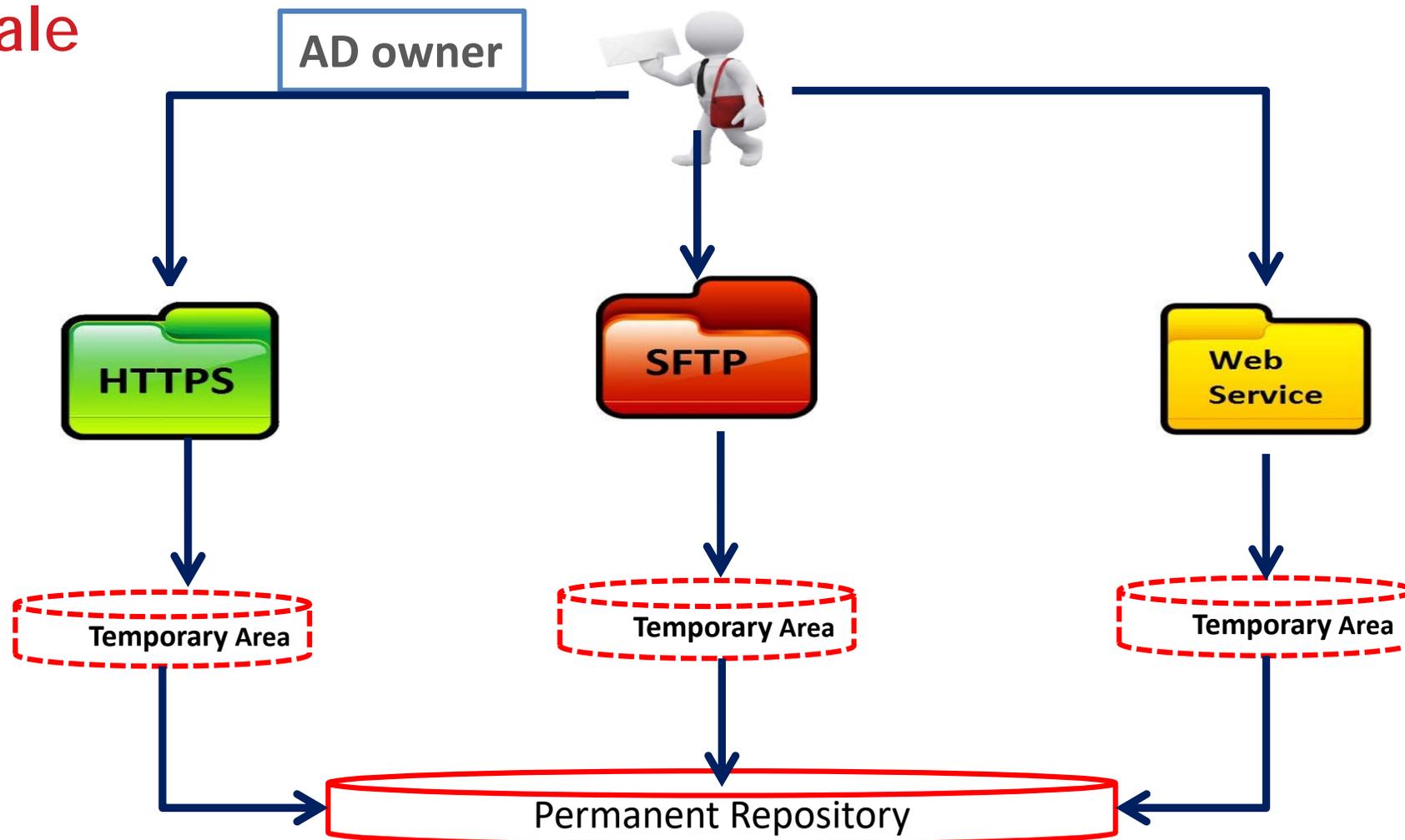
## L'esperienza dell'Istat (1)

### Acquisizione e trattamento dei dati amministrativi da parte dell'Istat

- ▶ **Infrastrutture:** apparati, processi e *governance* di elevata qualità per garantire l'**acquisizione** e il **trattamento** di grandi e complesse basi di dati amministrativi, tutelando la **riservatezza** delle informazioni.
- ▶ Acquisizioni nel 2018 attraverso il **Portale ARCAM:**
  - flussi da 50 enti (40 nel 2015);
  - 428 acquisizioni (280 nel 2015) a partire da 190 archivi amministrativi (90 nel 2015);
  - Agenzia delle Entrate, Inps e MIUR sono gli enti maggiormente coinvolti nella programmazione.

## L'esperienza dell'Istat (2)

### Il Portale Arcam



## L'esperienza dell'Istat (3)

### Framework per la verifica di qualità dei dati amministrativi acquisiti

IPERDIMENSIONE	DIMENSIONE
<b>FONTI</b>	F1. Informazioni sull'archivio
<i>Informazioni necessarie a gestire il processo di acquisizione dei dati con lo scopo di valutare e migliorare la qualità dei dati acquisiti</i>	F2. Rilevanza della fonte (utilizzi statistici)
	F3. Questioni connesse alle norme di uso (accessibilità)
	F4. Forniture acquisite e accordi
	F5. Relazioni e feedback con i titolari delle fonti
<b>METADATI</b>	M1. Chiarezza e interpretabilità dei dati
<i>Informazioni per la valutazione della qualità a livello concettuale e di processo presso il titolare della fonte</i>	M2. Comparabilità tra concetti amministrativi e concetti statistici (unità e variabili)
	M3. Descrizione del processo di acquisizione e trattamento (da parte del titolare)
<b>DATI</b>	D1. Technical checks
<i>Valutazione della qualità dei dati acquisiti</i>	D2. Integrabilità/Integrazione
	D3. Accuratezza e coerenza interna
	D4. Completezza
	D5. Aspetti temporali

## L'esperienza dell'Istat (4)

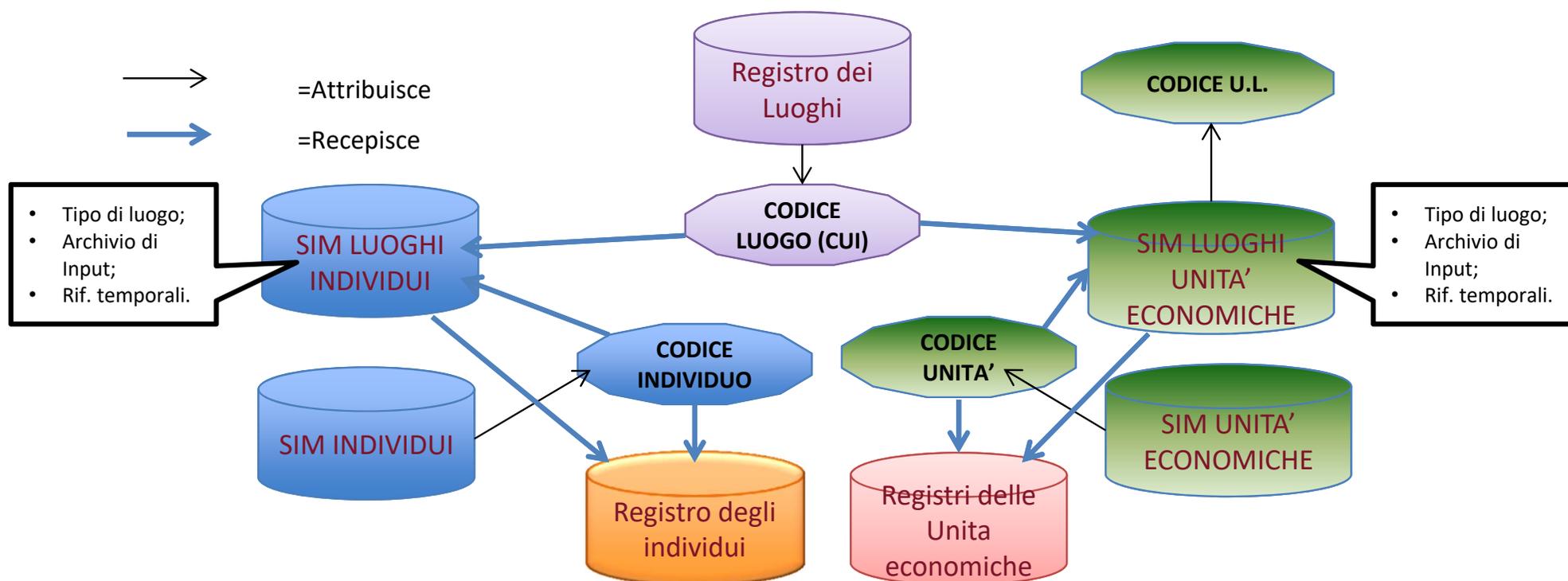
### SIM - Sistema di integrazione logico-fisica dei microdati

- ▶ **SIM:** dopo l'acquisizione attraverso ARCAM, i flussi di dati amministrativi vengono memorizzati in un repository di microdati integrati.
- ▶ Principali **finalità:**
  - creare un ambiente di elevata **sicurezza** per i dati amministrativi;
  - effettuare **controlli di qualità** sull'identificazione delle unità contenute negli archivi per garantirne la validità statistica;
  - attribuire un **codice** numerico "non parlante" alle unità statistiche;
  - procedere a **integrazioni di dati** e mettere a disposizione degli utilizzatori interni autorizzati gli archivi integrati, privi di dati identificativi diretti, necessari alla produzione dei dati statistici, nel rispetto della normativa in materia di **protezione dei dati personali**.

## L'esperienza dell'Istat (5)

### SIM - Sistema di integrazione logico-fisica dei microdati

Microdati riferiti a 3 tipi di unità statistiche: individui, unità economiche, luoghi associati



## L'esperienza dell'Istat (6)

### Il framework metodologico per l'utilizzo integrato delle fonti

- ▶ **Sistema integrato dei Registri (Sir):** ha l'obiettivo di realizzare Registri statistici (dati individuali integrati sull'intera popolazione di riferimento) in grado di produrre statistiche ufficiali attraverso l'utilizzo di più tipologie di fonti (amministrative, statistiche, Big data).
- ▶ **Funzioni:** gestione unitaria delle diverse tematiche (statistiche sociali, ambientali, economiche etc.) e integrazione concettuale e statistica tra le unità che lo compongono.
- ▶ **Componenti:** Registri statistici di base (Rsb), i Registri statistici estesi (Rse) e i Registri statistici tematici (Rst).
- ▶ **Output:** dati e indicatori statistici di maggiore qualità; elevata estensione/profondità informativa, utile a più finalità statistiche.

## I vantaggi del modello integrato: un esempio (1)

### 1) L'evoluzione delle statistiche su struttura/performance delle imprese

- ▶ **Vecchio modello:** due indagini dirette per misurare i risultati economici e occupazionali un universo di 4,4 mln di imprese. Risultati significativi solo a livelli molto aggregati di settore/dimensione/territorio; difficoltà di misurare la dinamica del sistema; elevato burden sui rispondenti.
- ▶ **Nuovo modello integrato (dati e metodologie innovative):**
  - **Dati amministrativi:** Studi di settore per 2,9 mln di imprese; bilanci per 700mila unità; imposte dirette per 600mila imprese; dati retributivi e contributivi Inps per 1,3 milioni di dipendenti; dati doganali per oltre 300mila imprese esportatrici/importatrici e altro ...
  - **Indagini dirette:** campionaria ridotta sulle piccole e medie imprese; totale per le grandi imprese, con forte abbattimento dei dati richiesti.

## I vantaggi del modello integrato: un esempio (2)

### Guadagni informativi:

- ▶ elevatissimo livello di **precisione delle stime** aggregate (le grandezze a livello macro ottenute per somma di dati individuali: 4,4 mln di imprese, con 17 mln di addetti, che generano oltre 700 mld di valore aggiunto)).
- ▶ **coerenza** tra analisi micro e analisi macro.
- ▶ possibilità di **analisi strutturali** delle eterogeneità all'interno del sistema delle imprese (territorio, settori, dimensioni, performance economica, internazionalizzazione, qualità del lavoro, retribuzioni ecc.).
- ▶ possibilità di **analisi dinamiche** con elevata valenza informativa/analitica (es.: decomposizione delle tendenze aggregate nei contributi delle singole unità/agggregazioni di unità ecc.).

## Prospettive: estensione del modello basato sui registri

- ▶ Progressiva **estensione** del modello basato sull'integrazione tra dati amministrativi e fonti statistiche a tutte le diverse aree statistiche: economiche, demografiche e sociali, ambientali e territoriali.
- ▶ **Censimenti permanenti**: piena realizzazione delle potenzialità derivanti dall'uso integrato dei dati amministrativi a fini statistici:
  - Produzione di **dati censuari a cadenza annuale/pluriennale** e non più solo decennale come nel vecchio modello.
  - **Ingenti risparmi**: economie, rispetto alla spesa sostenuta nella precedente tornata di svolgimento dei censimenti decennali di tipo tradizionale (anni 2010 e 2011), di oltre 300 milioni di euro, pari a circa il 49%, nel round censuario 2018-2021 e, a regime, di circa 370 milioni di euro, pari al 58%, nel round censuario del 2022-2031.

**Grazie per l'attenzione**