

Seminario
URBES, ARCHIMEDE, Censimento permanente
I Comuni verso l'uso statistico degli archivi amministrativi
e dei sistemi di integrazione delle fonti
18 giugno 2015, Terni

Il progetto ARCH.I.M.E.DE
integrazione centralizzata e generalizzata di
fonti amministrative

Obiettivo

«..ampliamento dell'offerta informativa dell'ISTAT mediante produzione di collezioni di dati elementari da rendere disponibili all'utenza, utili alla ricerca sociale ed economica, alla programmazione territoriale e settoriale, alla valutazione delle politiche pubbliche a livello nazionale, regionale e locale.»

(Delibera 12/ DGEN del 18/01/2013)

Metodo

Sfruttamento dei contenuti informativi
presenti nel S.I.M.

(Sistema Informativo per l'Integrazione di fonti amministrative)

! contenuto informativo \neq variabile

! presente \neq evidente

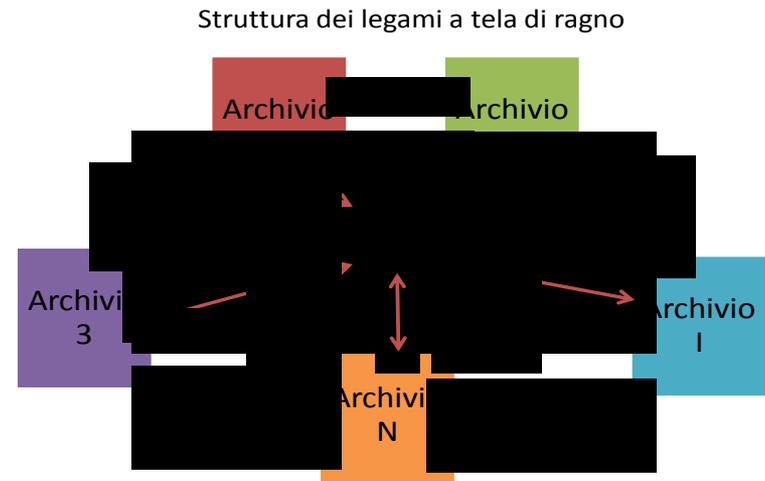
Repository dei dati amministrativi acquisiti dall'Istat creato allo scopo di:

- supportare i processi di produzione statistica dell'Istat,
- favorire l'utilizzo di dati individuali privandoli degli identificativi diretti,
- aumentare le potenzialità informative derivanti dal processo di integrazione

Es.

- realizzare e aggiornare registri statistici,
- sottosistemi di informazioni statistiche dedicate a lavoro, mobilità, istruzione, reddito, ecc.,
- stratificare campioni o fornire variabili ausiliare o integrative

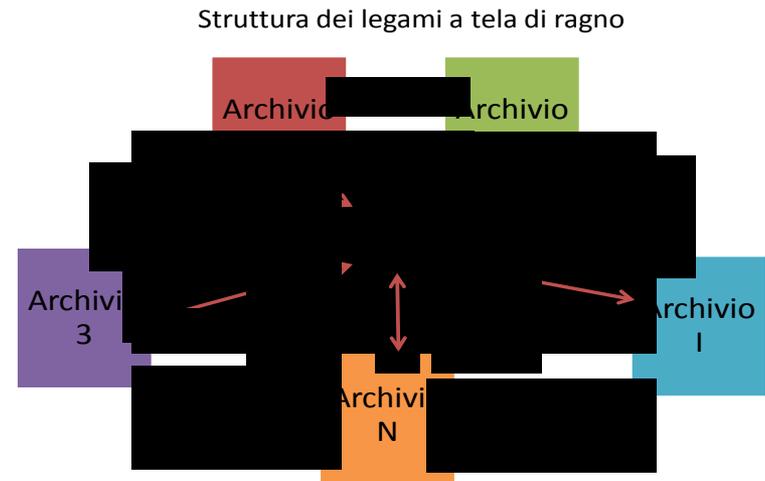
E' una infrastruttura statistica di base
nella quale vengono **RELAZIONATI**
n archivi di fonte amministrativa
riferiti agli *oggetti*: individui,
famiglie e unità economiche



S.I.M. Sistema Integrato di Microdati

Tipologia delle fonti		Sottosistemi	
		Individui	Unità
Anagrafici	Anagrafi Comunali	X	
	Anagrafi Consolari	X	
	Anagrafe Tributaria	X	X
	Permessi di soggiorno	X	
Fiscali	Banca Dati Reddittuale - MEF	X	
	Studi di settore		X
	Modello UNICO	X	X
	Modello 730	X	
	Modello 770	X	X
Formazione	Anagrafe degli studenti	X	
	Anagrafe degli studenti universitari	X	
	Anagrafe personale doc. e non doc. delle scuole	X	X
	Anagrafe personale doc. e non doc. delle università	X	X
Lavoro	Arch. INPS Emens (UNIMENS)/Parasubordinati/Cassa integrazione/Lav. agricoltura/Artigiani e commercianti/ Autonomi dell'agr./Lav Domestici	X	X
	Archivi INAIL	X	X
	Arch. Ex-INPDAP ed Ex-ENPALS	X	X
	Cedolini stipendiali (MEF)	X	X
	Casellario dei pensionati	X	
Welfare	ANF/Maternità	X	
	Mobilità/Disoccupazione/LSU	X	
Camerali	Registro delle Imprese		X
	Soci delle Imprese	X	X
	Persone con cariche sociali	X	X
	Bilanci delle Imprese		X

Ad ogni oggetto è associato in ogni fonte lo stesso **ID univoco e stabile** nel tempo. TRA gli oggetti sono evidenziate alcune delle **relazioni logiche e fisiche** che esistono, e fissate nel tempo t e nello spazio s



Caratterizzati da fonti amministrative di riferimento specifiche

S.I.M. Individui (42 fonti)

Alimentato da tutte le fonti in cui sono contenuti dati sufficienti per l'identificazione degli individui

Obiettivo: consentire agli utilizzatori di disporre del più ampio set possibile di unità elementari

S.I.M. Unità economiche (42)

Alimentato da es. fonti amministrative che permettono di intercettare i legami tra partite IVA e codici fiscali, ecc.

! Ogni annualità della stessa fonte conta una unità

S.I.M. luoghi unità economiche (30 fonti)

Repository delle localizzazioni delle unità economiche presenti nelle fonti amministrative.

S.I.M. luoghi individui (25 fonti)

Repository dei luoghi che in qualche modo possono interessare le persone fisiche riconosciute dai data set amministrativi inseriti nel sistema: residenza anagrafica, domicilio fiscale, luoghi di lavoro, luoghi di studio,

S.I.M. relazioni tra unità economiche (7 fonti)

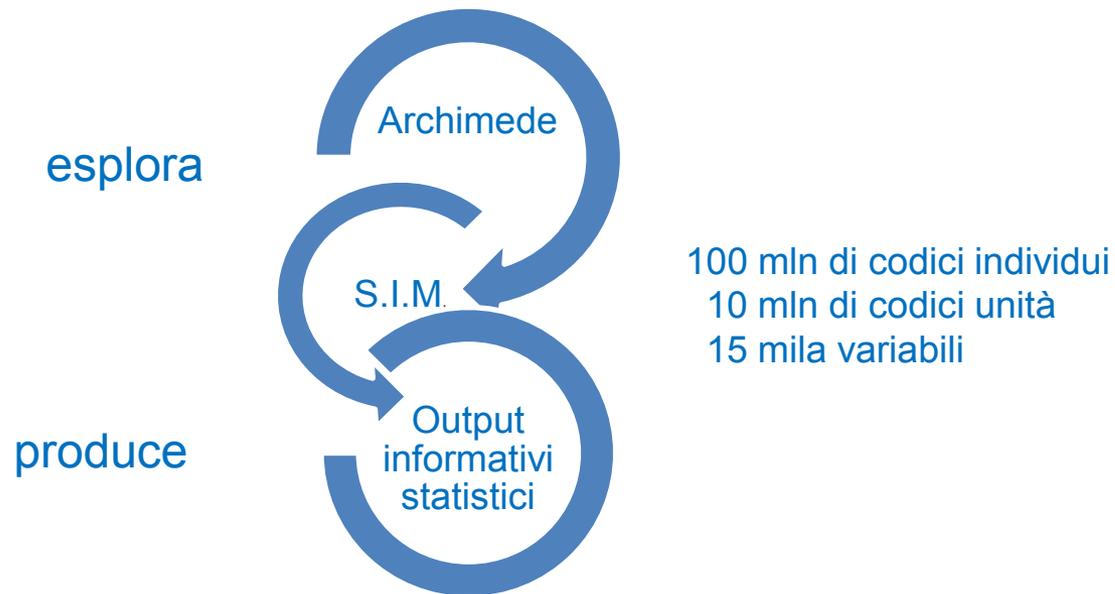
Obiettivo: cogliere alcune possibili relazioni tra unità economiche quali, ad esempio, eventi di trasformazione e legami societari.

S.I.M. relazioni tra individui (3 fonti)

Obiettivo: identificare le relazioni tra gli individui; famiglia anagrafica e famiglia “fiscale”.

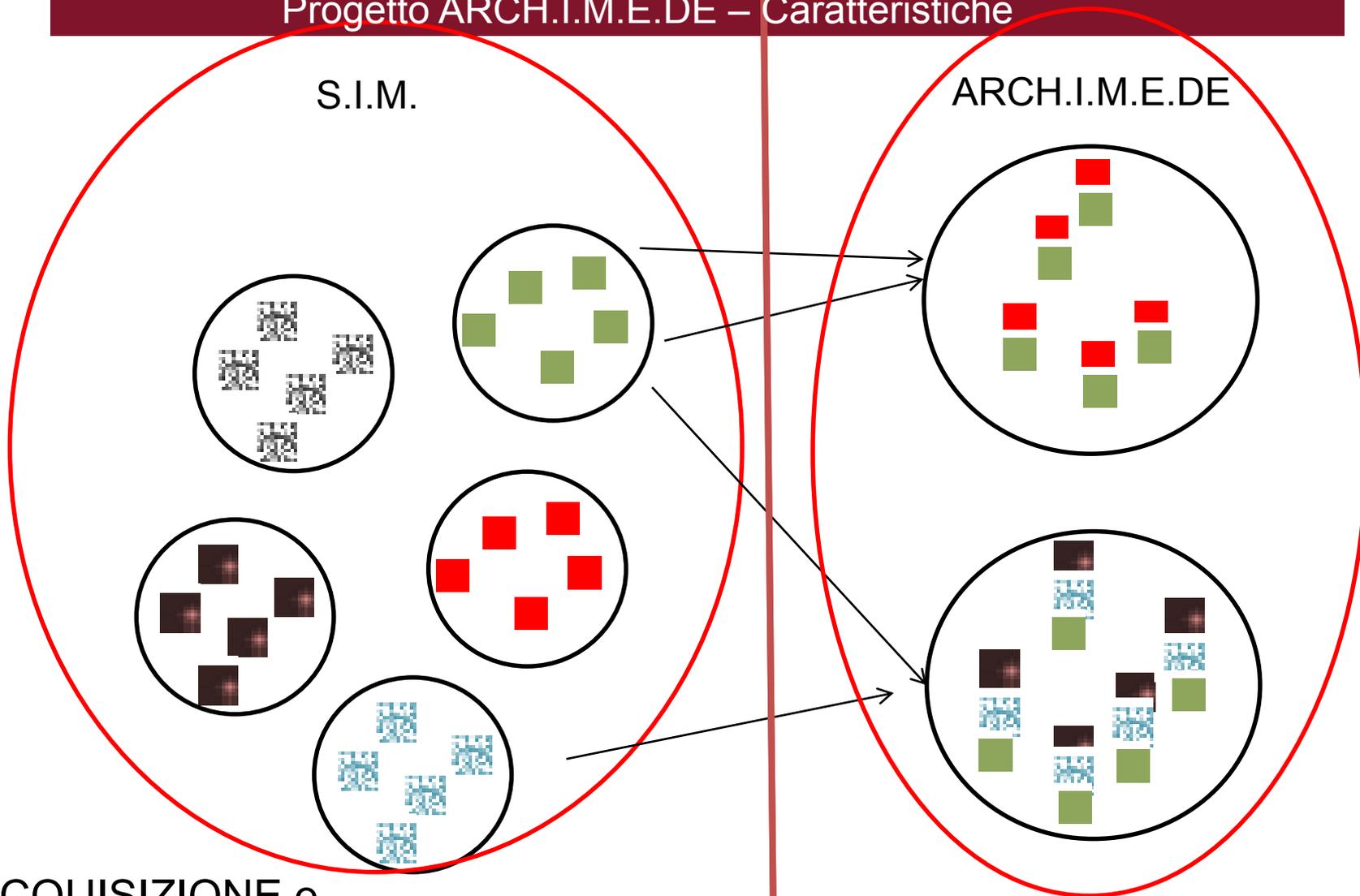
S.I.M. relazioni tra individui e unità economiche (12 fonti)

Obiettivo: integrare le informazioni sulle relazioni tra gli individui e le unità economiche. Presenza contestuale dei due identificativi necessari per legare i domini “socio-demografico” e “economico”: codice individuo e codice unità.



S.I.M.

ARCH.I.M.E.DE



ACQUISIZIONE e
INTEGRAZIONE
LOGICO/FISICA

INTEGRAZIONE
INFORMATIVA

La **disponibilità dell'informazione** è ENORME

Identificare l'informazione statistica che può soddisfare un bisogno (*data mining*)



Esplorare ?

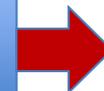
Le **definizioni** e le **classificazioni** possono essere determinate all'interno del processo di esplorazione e quindi non fissate a priori (*outputs* non strettamente vincolati ai sistemi di regolamentazione europei)

L'adeguatezza dell'informazione prodotta rispetto agli schemi concettuali statistici può essere valutata solo ex-post

USO DEI DATI AMMINISTRATIVI tradizionale

Uso “Diretto” delle Fonti Amministrative

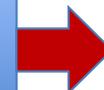
- Realizzazione di Registri Statistici
- Stime di variabili per domini di interesse



**SOSTITUZIONE
della Ind. STAT:**

Uso “Indiretto” delle Fonti Amministrative

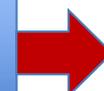
- Miglioramento di processi di produzione (disegno campionario)
- Informazione ausiliaria nella fase di controllo dei dati



**SUPPORTO
alla Ind. STAT:**

USO DEI DATI AMMINISTRATIVI NEL PROGETTO ARCHIMEDE

ESPLORAZIONE dell'informazione disponibile
Ampliamento dell'offerta informativa



**PRODUZIONE
di nuove
“tipologie” di
informazione
statistica**

Output informativi statistici:

Collezioni di microdati integrati
e
Collezioni di macrodati



Utenti interni

Utenti esterni



Documentate
qualità
Contenuti

Replicabili
(industrializzati)

Territoriali
Informazioni geo referenziate
Tempo
Informazioni longitudinali

- **Eleva** il livello di autonomia degli utenti interni ed esterni
- **Omogeneizza** i risultati degli utenti

Output informativi statistici:

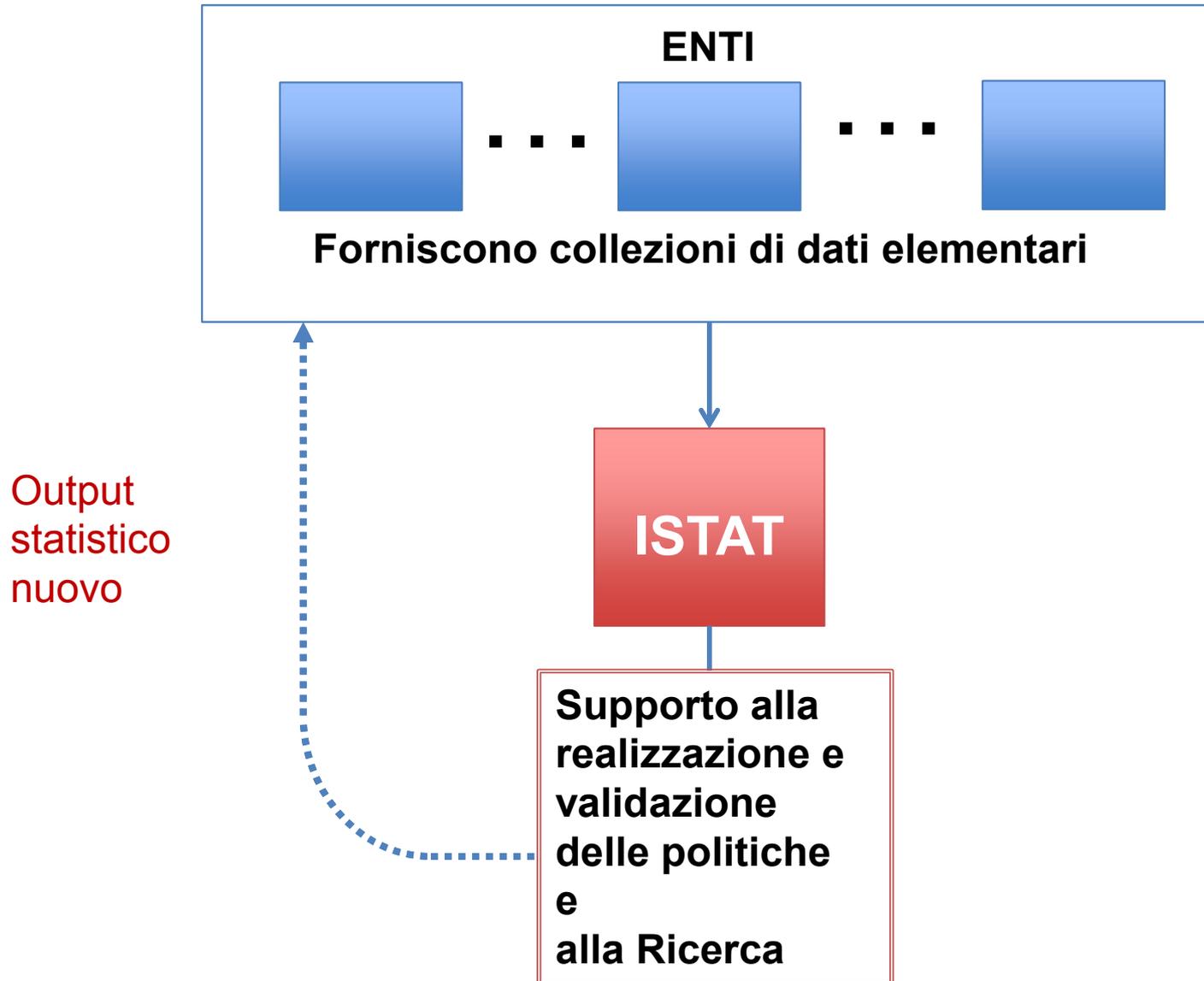


On demand: realizzati sulla base di specifiche esigenze informative di utenti

Generalizzati: realizzazione di specifiche offerte informative identificate sulla base delle potenzialità informative di SIM



- Fare riferimento ad un insieme di unità che possono non rappresentare l'universo di una specifica popolazione
- Utilizzare dati amministrativi di SIM anche non trattati statisticamente
- Integrare anche basi informative fornite da utenti esterni (es. Comuni/Regioni)
- Utilizzare concetti non necessariamente coerenti con la statistica ufficiale (es. Regolamenti europei)



2013 : Attività sperimentale

Popolazioni che insistono su di un territorio

Precarietà lavorativa

Condizioni Socio-economiche delle famiglie

2014

- Realizzazione di una proposta di indicatori derivati dalle basi sperimentali (Ambito GdL 5 del Comitato Scientifico del Censimento)

- Nuove sperimentazioni (*concluse*):
 - Bacini Universitari
 - P.IVA Monocommittenti
 - Legami familiari tra individui non coabitanti

- Nuove sperimentazioni (*in atto*):
 - Utilizzo Big Data per la stima della differenti tipologie di *City users*
 - Metodi di classificazione delle famiglie (Data Mining / Indici sintetici)

2015

1. Definizione del «cruscotto» degli **indicatori** proposti (*Febbraio*)
2. Termine della sperimentazione dei partecipanti e successiva valutazione dei risultati (?).
3. Inizio sperimentazione progetto su «percorsi di istruzione/formazione/inserimento lavorativo» con Lombardia/Lazio/Veneto/Em. Rom./Sicilia/Valle d'Aosta.
4. Inizio del progetto sperimentale a supporto della Protezione Civile sulla quantificazione della popolazione in zone di rischio (Giugno)
5. Test per il Passaggio del sistema P&P da livello comunale a sub-comunale per le aree metropolitane e i comuni di maggiori dimensioni
6. Revisione delle metodologie di calcolo del reddito individuale da fonti amministrative
7. Produzione e diffusione, a regime, delle basi dati a partire dall'anno di rif. 2013 (*Entro fine anno*)

PSN 2014-2016

- sperimentazione sulle tre basi realizzate da parte degli *uffici di statistica compartecipanti* :
 - Comuni di Firenze, Bologna, Brescia, Trieste, Modena, Milano
 - Regioni e Prov. Autonome
- predisposizione bozza nota informativa ai compartecipanti
- trasmissione dei microdati successiva alla pubblic. del PSN
- risultati attesi entro sei mesi

Aggiornamento PSN 2015-2016

- Diffusione micro/macro delle tre basi di dati
- Nuova sperimentazione su: Percorsi di istruzione, formazione, abbandono scolastico e inserimento lavorativo degli individui tra i 14-29 anni

Aggiornamento PSN 2016

- I quattro progetti sono stati separati in schede differenti con compartecipazione di tutti gli Uff. di statistica di Comuni, Province, Regioni
- Inserimento di una nuova scheda su Identificazione delle popolazioni in territori a rischio (Collaborazione con Dipartimento della Protezione Civile)

Nota informativa per gli enti compartecipanti

Gli Uffici di statistica, in relazione alle rispettive finalità istituzionali e agli ambiti territoriali di competenza, hanno il compito di valutare le basi di microdati prodotte dall'Istat indicate al paragrafo 1 con riferimento ai seguenti aspetti generali:

- completezza, congruità e utilità del contenuto informativo delle basi di microdati rispetto alle esigenze conoscitive dello specifico fenomeno oggetto di studio connesse alle finalità istituzionali dell'ente di appartenenza;
- vantaggi derivanti all'Ufficio di statistica in termini di risparmio di costi e di miglioramento della qualità dell'informazione statistica prodotta.

La valutazione di ciascuna base di microdati è effettuata con le modalità di seguito precisate:

Base di microdati relativa a:

1.1 Oggetto di studio 1:

- Definizione dell'ambito di ricerca e degli obiettivi specifici dell'oggetto di studio:
- Specifiche attività da svolgere:
- Output sperimentale previsto:



L'Istat fornirà le basi di microdati entro 15 giorni dalla data di pubblicazione del PSN in Gazzetta Ufficiale

I risultati delle valutazioni effettuate dovranno essere trasmessi all'Istituto entro sei mesi dalla data di fornitura delle basi di microdati.

- ***Verso l'esterno*** dell'Istituto:
 - Diffusione microdati ad un livello di aggregazione territoriale fine
 - Messa a disposizione di Enti Pubblici di basi di microdati per l'analisi di sottopopolazioni di interesse (schematizzazione delle popolazioni a cui sono rivolti interventi e politiche pubbliche)
 - Possibilità di identificare strumenti omogenei – indicatori – fra le varie realtà territoriali a supporto alla comparazione territoriale

- ***Verso l'interno*** dell'Istituto:
 - Esplorazione/identificazione di segnali delle fonti amministrative utili ai processi statistici.
 - Ponte fra statistiche sulle imprese e statistiche sugli individui/famiglie
 - Sperimentazione di nuove metodologie di integrazione e di analisi
 - Possibilità di analizzare fenomeni per «popolazioni» diverse

Tempi di disponibilità delle fonti integrate

Possibilità di sperimentazioni sulle fonti

Distanza con output statistici «simili»

Valutazione della qualità

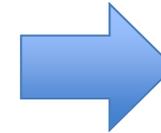
Criteri / Strumenti di diffusione

Uso sperimentale dei dati amministrativi:

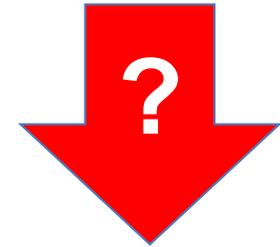
- **Archimede**
- **ISTAT**
- **SISTAN**



Sperimentazioni
congiunte



VINCOLO
PSN



Identificazione di nuovi strumenti, **differenti dal PSN**, che, nel rispetto delle norme vigenti relativamente alla riversatezza e al vincolo dell'“informativa” permettano con la massima trasparenza ai ricercatori dell'Istat e del Sistan di sviluppare sperimentazioni su dati integrati di fonti amministrative e di indagine

Alcune specificità del progetto

- fare riferimento ad intere sottopopolazioni (e non quindi a campioni statistici di popolazioni);
- essere riferite ad ambiti territoriali ristretti (piccoli comuni o ambiti sub comunali);
- poter contenere una molteplicità di informazioni (studio+lavoro+tipologia familiare+reddito....);
- poter contenere dati sensibili (ad es. contributi per disabilità);
- poter derivare sistemi di classificazione e definizione non “omogenei” con quelli della statistica ufficiale.

**Necessità di progettare un ambiente di diffusione
«autonomo» ?**

**Necessità di pensare la diffusione dei microdati –
anche ad enti SISTAN – in maniera differente (senza
identificativi diretti) ?**

Grazie per l'attenzione

Roberta Vivio vivio@istat.it