

Produzione, diffusione e utilizzo di microdati: esperienze, progetti e prospettive di integrazione ai fini della ricerca e del disegno delle politiche pubbliche

Giorgio Alleva

Presidente dell'Istituto Nazionale di Statistica

L'integrazione tra archivi amministrativi e indagini campionarie come strumento per le analisi socio-economiche, il disegno e la valutazione di politiche pubbliche

Fondazione Bruno Kessler

17 ottobre 2016

Outline

Lo scenario

- Una nuova domanda
- I microdati
- Il processo di modernizzazione

Esperienze tra passato e futuro

- Demografia
- Lavoro
- Famiglie e imprese

Prospettive

- Laboratori virtuali

Conclusioni

Lo scenario

Una nuova domanda

Mentre il singolo individuo è un enigma irrisolvibile, quando è insieme agli altri diviene una certezza matematica. È impossibile, per esempio, predire il modo in cui agirà un uomo, mentre è invece possibile dire con precisione cosa faranno un certo numero di uomini messi insieme. L'individuo varia, ma le percentuali rimangono costanti. Così dicono le statistiche.

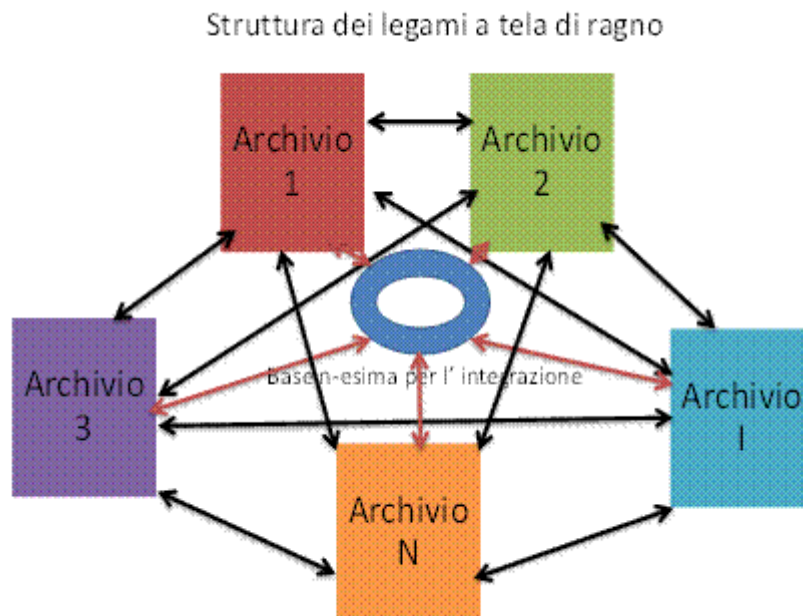
(William Winwood Reade citato da Sherlock Holmes in «Il segno dei quattro» di Arthur Conan Doyle)

Forte incremento della domanda di microdati, da parte delle istituzioni e della comunità scientifica, per analisi a carattere tematico (economico, sociale, ambientale, ecc.) e territoriale.

- Il microdato [dato elementare riferito alla singola unità statistica] diventa strumento essenziale per la valutazione dell'impatto delle politiche.
- La gestione dei microdati consente all'Istat di rispondere alle **diverse esigenze istituzionali**, ma richiede diverse attenzioni :
 - esigenze degli utilizzatori ;
 - protezione dei dati personali;
 - sicurezza.

- I microdati prodotti, sono archiviati nell'ARchivio dei MIcroDATi di Istituto (**ARMIDA**) e resi disponibili tramite il Laboratorio ADELE o il *Contact centre* .
- Dal 2013 disponibili **File ad uso pubblico**, collezioni di dati elementari scaricabili liberamente e gratuitamente dal sito Istat (micro.STAT).

SIM: Sistema Integrato di Microdati



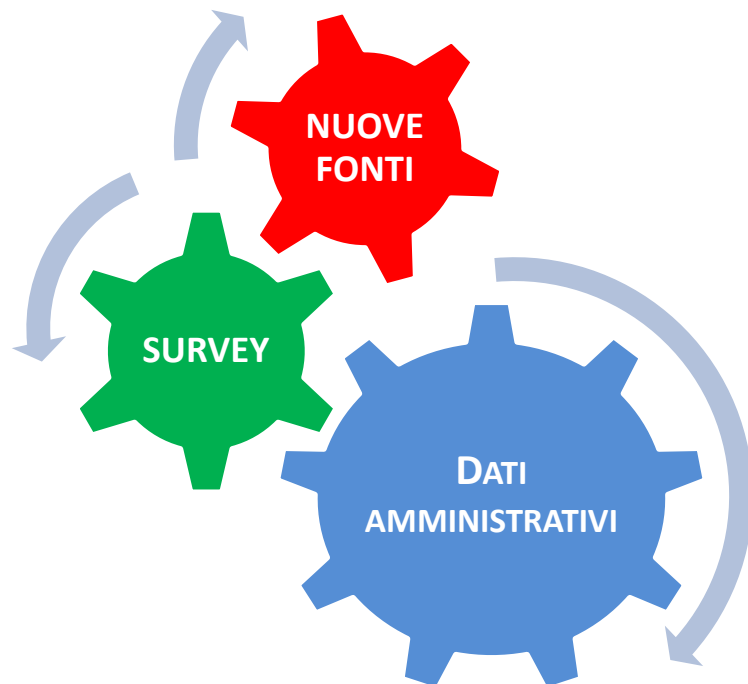
SIM è una infrastruttura statistica di base, già esistente, nella quale vengono integrati dati riferiti a individui, famiglie e unità provenienti da una pluralità di basi dati di fonte amministrativa

Il processo di modernizzazione dell'Istat

- Le **indagini campionarie** sono divenute sempre più costose. I tassi di risposta sono in diminuzione ed è necessario contenere il fastidio statistico sui rispondenti.
- La digitalizzazione dei **dati amministrativi** e il miglioramento delle tecniche di matching fra i dati permettono di fare maggiore ricorso alle fonti amministrative.
- Le **nuove tecnologie** hanno aumentato la capacità di archiviare, processare e analizzare quantità sempre maggiori di dati.

Il processo di modernizzazione dell'Istat

- La capacità di **estrarre valore dai dati** è legata alla capacità di **integrare** dati che provengono da fonti differenti.
- Mettere a confronto fonti differenti garantisce guadagni in termini di **accuratezza, coerenza, completezza** delle informazioni statistiche prodotte.



Si tratta di un **percorso intrapreso** da molti istituti di statistica dei paesi avanzati.

Il processo di modernizzazione dell'Istat

- Modello basato sull'utilizzo dei **registri statistici**, derivati dalle fonti amministrative ed eventualmente integrati con indagini.



I registri di base

- L'utilizzo di **dati amministrativi** e la loro integrazione permetterà di:
 - aumentare il dettaglio di analisi;
 - mettere insieme i percorsi sociali ed economici di individui e imprese ("scrivere" le storie individuali);
 - connettere a livello micro i fenomeni economici e sociali;
 - descrizione dei fenomeni a livello micro-territoriale.
- Le **indagini** continueranno ad essere utilizzate per completare il quadro informativo, analizzare fenomeni specifici, fornire risposte a determinate chiavi di lettura, individuare nuovi trend.
- Nel futuro i **big data** saranno utili per ampliare le opportunità di analisi, aumentare la tempestività delle informazioni, contribuire a migliorare la qualità delle stime.



Ad esempio...

Il censimento permanente

- Attualmente l'Istat conduce numerose indagini sociali di alta qualità (mercato del lavoro, redditi e condizioni di vita, il sistema multiscopo, indagine sulle spese delle famiglie) che costituiranno un sistema integrato.
- Necessità di **integrazione e armonizzazione** per ridurre il carico statistico, per migliorare la qualità delle informazioni e la loro confrontabilità al fine di garantire lo studio delle interrelazioni dei fenomeni.
- **Integrazione** delle indagini sociali nel **Censimento permanente**.

Esperienze tra passato e futuro

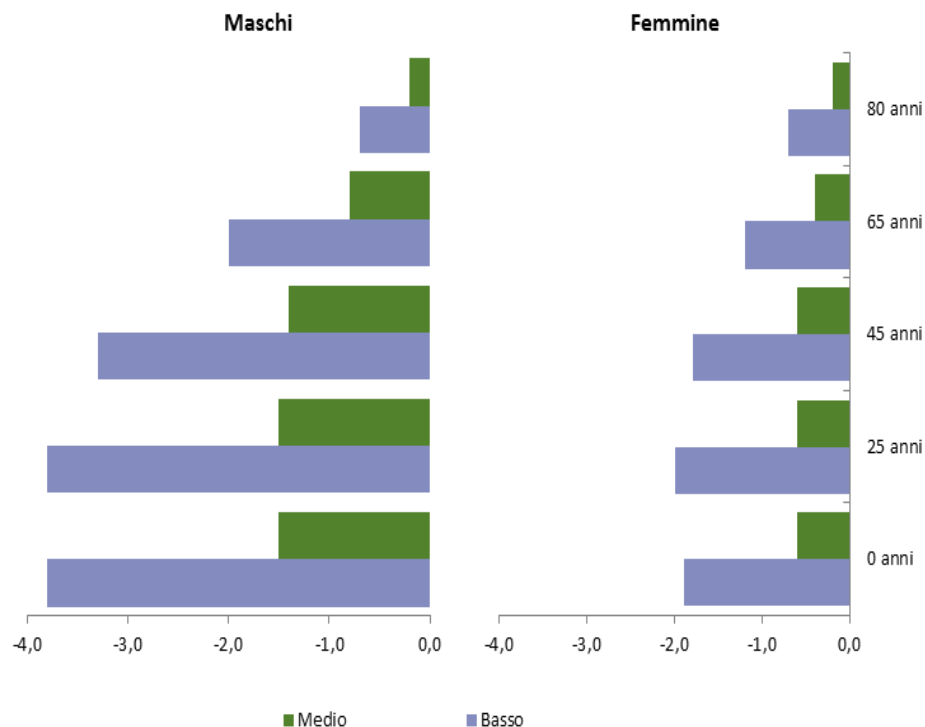
Demografia: fonti anagrafiche ed amministrative

- Inizialmente il registro degli individui sarà alimentato da dati di fonte anagrafica che popoleranno **l'Anagrafe Virtuale Statistica** (ANVIS). Successivamente l'integrazione con altre fonti amministrative consentirà una **lettura micro-longitudinale** di tutti i comportamenti e le relazioni e delle loro rispettive mutazioni.
- **Esempi:**
 - Archivio longitudinale dei fratelli per l'analisi della fecondità
 - Ricostruzione della storia della popolazione con background straniero (acquisizione della cittadinanza)
 - Anagrafe degli studenti per l'analisi longitudinale dei percorsi educativi
 - Archivi INPS e Ministero del lavoro e politiche sociali per l'analisi delle carriere lavorative, dei flussi di entrata e uscita dal mercato del lavoro

Un esempio di integrazione tra fonti

DIFFERENZA IN ANNI DELLA SPERANZA DI VITA RISPETTO A LIVELLO DI ISTRUZIONE ALTO

Differenze in anni. Anno 2012



L'effetto del titolo di studio sulla speranza di vita è più marcato tra gli uomini.

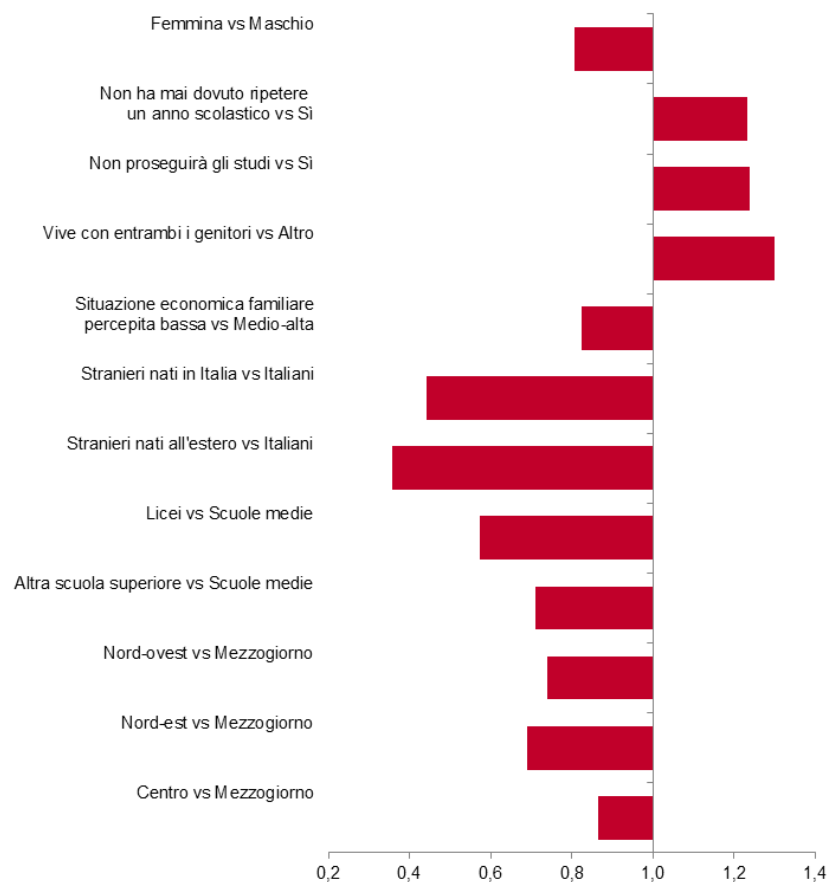
Ad esempio si osserva che un uomo di 25 anni con titolo di studio basso vive **3,8** anni meno di un coetaneo con livello di istruzione elevato

FONTE: Istat, Indagine su decessi e cause di morte, anno 2012 e Censimento della popolazione e delle abitazioni, anno 2011.

Un esempio di analisi causale su fenomeni emergenti

STIMA DELLA PROBABILITÀ DI VOLER VIVERE IN ITALIA DA GRANDI, STUDENTI STRANIERI E ITALIANI

Rapporti di probabilità. Anno 2015



Meno propensi a voler restare in Italia:

- * le ragazze
- * gli stranieri in particolare se nati all'estero

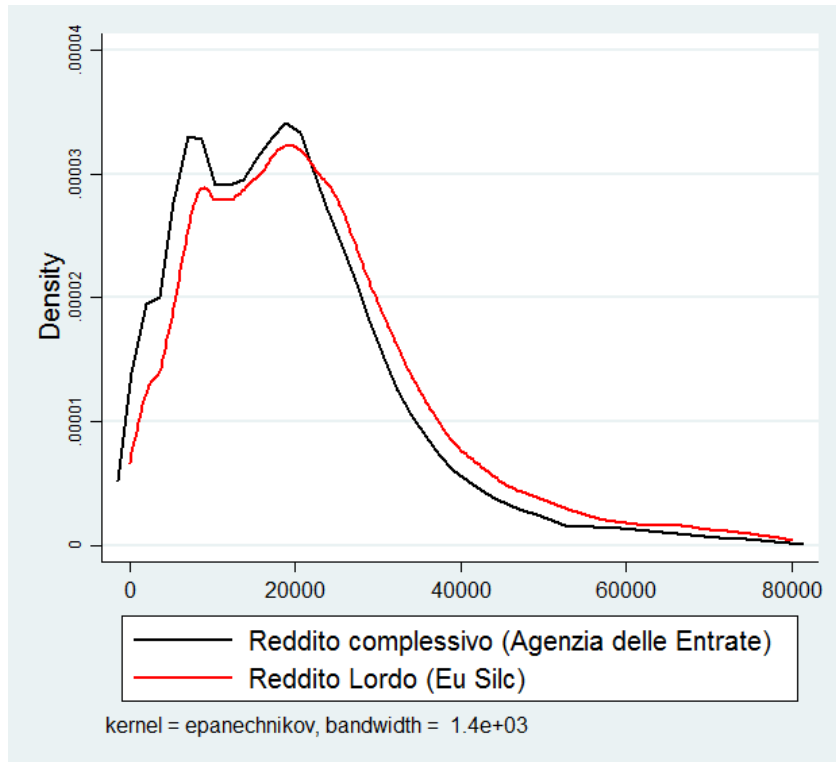
FONTE: Istat, Indagine sull'integrazione delle seconde generazioni.

- Crescente domanda di informazioni per le statistiche sul lavoro:
 - Quali sono le componenti in entrata e uscita dall'occupazione che determinano le variazioni degli stock rilevate?
 - La ripresa occupazionale è transitoria o strutturale?
 - Quanto è concentrata o diffusa, settorialmente e territorialmente, la capacità di creare posti di lavoro?
 - Qual è lo stato e come si evolve il *mismatch* tra formazione dei lavoratori e domanda di qualificazione da parte delle unità economiche?

- **Accordo** siglato tra il Ministero del Lavoro, l'Istat, l'Inps e l'Inail; l'obiettivo è la realizzazione e diffusione di prodotti congiunti che facilitino la lettura integrata delle informazioni riguardanti il mercato del lavoro.
- Output congiunti richiedono un approccio innovativo rispetto al passato:
 - **combinare** le informazioni statistiche;
 - fornire una **visione unitaria** dei fenomeni anche attraverso la diffusione di microdati per la ricerca;
 - **utilizzare ed integrare** dati provenienti da istituzioni diverse;
 - adottare una comunicazione che segua la **"logica ad albero"**.

Famiglie: integrazione di fonti

- EU-SILC come base dati nei modelli di microsimulazione.



- EU-SILC è la fonte di dati di riferimento per il modello di microsimulazione **FaMiMod**.
- La versione italiana di EU-SILC, utilizza **microdati sia campionari sia amministrativi**.
- Altro esempio di integrazione di fonti è **AD-SILC** che associa le informazioni relative alla storia lavorativa degli individui a variabili socio-demografiche.

Famiglie: un esempio di valutazione delle policy

Tavola 5 - Reddito di "cittadinanza" (ddl. n. 1148). Indicatori di disuguaglianza e di povertà - Anno 2015

DISUGUAGLIANZA	Prima	Dopo
Gini	0,300	0,282
POVERTÀ	Prima	Dopo
Incidenza della povertà relativa	17,395	17,395
Povertà grave (50% della L. di Povertà)	3,898	0,037
Poverty gap ratio (a)	5,707	2,870

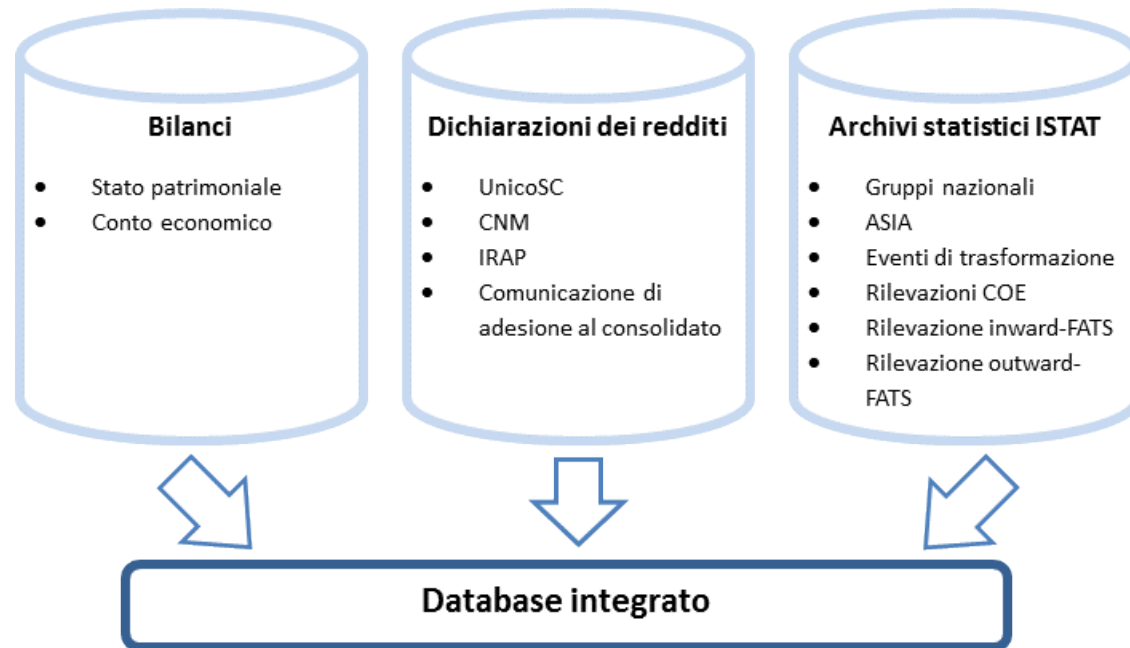
Fonte: Istat, Modello di microsimulazione sulle famiglie

(a) Differenza percentuale tra il reddito medio dei poveri e la linea di povertà.

- Il modello **FaMiMod** è stato utilizzato per valutare la proposta di inserimento del «reddito di cittadinanza»
- In futuro ulteriori analisi saranno possibili grazie all'integrazione dei dati **EU-SILC** con i **consumi** e **la ricchezza** delle famiglie.

Imprese: integrazione di fonti

- Modello di microsimulazione **ISTAT-Matis** (Modello per l'analisi della tassazione e degli incentivi sulle società di capitali).

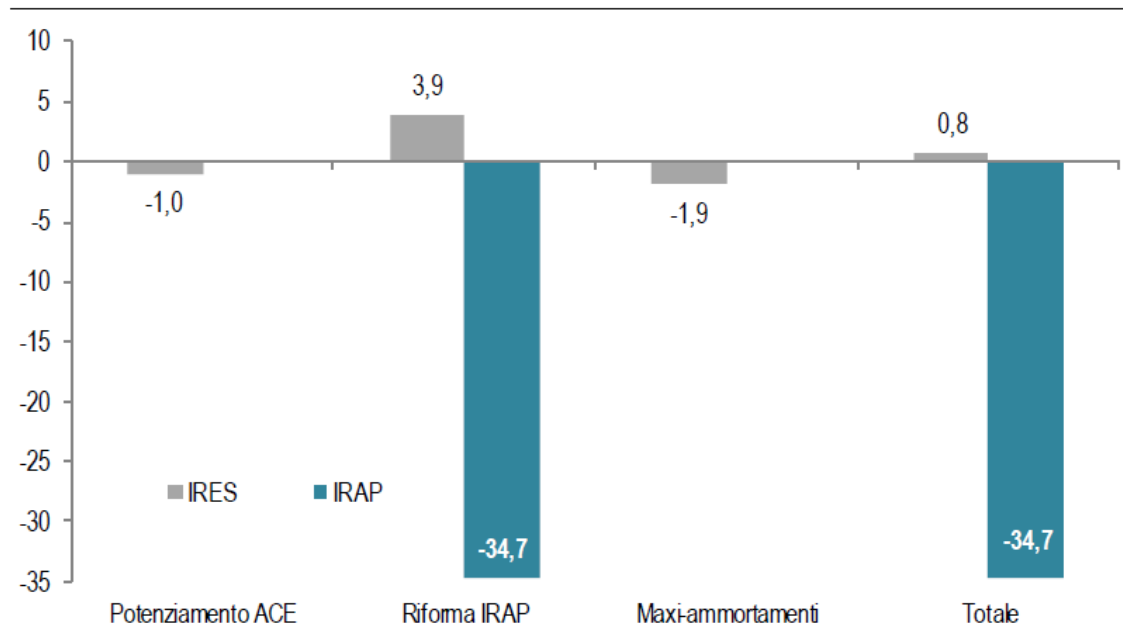


- Il modello si basa su un database che **integra** tra loro dati fiscali, amministrativi e di indagini statistiche .

Imprese: un esempio di valutazione delle policy

- Il modello **ISTAT-Matis** è stato utilizzato per analizzare gli effetti dei provvedimenti in materia di tassazione dei redditi delle società di capitali in vigore nel biennio 2015 - 2016.

FIGURA 1. EFFETTI DEI PROVVEDIMENTI SUL PRELIEVO IRES E IRAP RISPETTO ALLA NORMATIVA VIGENTE AL 2014. Anno di imposta 2016, punti percentuali



Prospettive

Il regolamento europeo sulle indagini sociali

- È in corso di approvazione il regolamento europeo sulle **indagini sociali** che definisce le dimensioni sociali che ogni Istituto nazionale di statistica è tenuto a rilevare in **forma armonizzata** a livello europeo:
 - ✓ Mercato del lavoro
 - ✓ Reddito e condizioni di vita
 - ✓ Salute
 - ✓ Istruzione e formazione
 - ✓ Uso ICT
 - ✓ Uso del tempo
 - ✓ Consumo

I microdati e la recente normativa

La maggiore disponibilità di microdati integrati garantirà **nuove opportunità di ricerca**.

La recente normativa (D.lgs 97/2016) semplifica l'accesso ai microdati:

- **superando** alcune criticità emerse nel passato:
 - definizione più ampia di “**ente di ricerca**”;
 - disponibilità di microdati relativi alle **persone giuridiche**.
- **aprendo** nuove prospettive:
 - costituzione di **laboratori virtuali**.



- Il nuovo quadro normativo demanda al **Comstat** il compito di definire **le linee guida** per consentire l'accesso ai dati.
- Le linee guida dovranno individuare:
 - criteri per il **riconoscimento degli enti**;
 - criteri di **ammissibilità** dei progetti di ricerca;
 - **organizzazione dei laboratori** e criteri per l'accertamento **dell'idoneità** dei soggetti per la gestione dei laboratori virtuali;
 - **sanzioni** in caso di violazione degli impegni assunti.

Conclusioni

L'agenda statistica del futuro

- Il microdato darà ai ricercatori e policy makers l'opportunità di studiare relazioni più complesse, verificare l'impatto delle politiche, analizzare l'evoluzione dei fenomeni economici, sociali e demografici.
- Vantaggi:
 - dall'approccio macro/trasversale all'approccio **micro/longitudinale**;
 - realizzazione di **prodotti integrati** in grado di mettere in connessione dimensioni diverse.
- **Nuove metodologie** e **nuove competenze** sono necessarie.