

# Accesso ai microdati, ricerca scientifica e valutazione delle politiche: urge un cambio di passo

Categories : [Istituzioni e regole](#)

Tagged as : [Menabò n. 110/2019](#), [Ugo Trivellato](#)

Date : 1 Ottobre 2019

Una premessa torna utile. Mi riferirò alla ‘ricerca scientifica’ comprendendovi anche il ‘disegno e valutazione di politiche pubbliche’ e la ‘statistica’, perché i tre ambiti, pur diversi, sono assimilabili per il trattamento di microdati. Infatti, tutti sono interessati alle informazioni individuali non di per sé, ma nella prospettiva di caratterizzare fenomeni collettivi – leggi di comportamento, relazioni di causalità, pattern di fenomeni –, che trascendono gli individui (Trivellato, 2019, pp. 27-30; salvo diversa indicazione, tutti i riferimenti bibliografici e alla normativa sono in [Data-Driven Impact Policy Evaluation](#)).

## L’evoluzione dei rapporti fra enti produttori di dati e comunità scientifica.

La rivoluzione digitale investe la produzione e l’utilizzo dei microdati alla soglia degli anni ’70. A fronte della crescita impetuosa del processo di digitalizzazione, che tocca pressoché ogni aspetto della vicenda umana e si impone a livello globale, lo schermo del segreto statistico e delle regole deontologiche proprio degli istituti nazionali di statistica (INS) e dei ricercatori si rivela ben presto inadeguato. Nell’Unione Europea (UE) l’esito è l’adozione della Direttiva [19]95/46/CE “*relativa alla tutela delle persone fisiche con riguardo al trattamento dei dati personali*”. (Si ricorda che le Direttive vincolano gli Stati membri rispetto al risultato da raggiungere, ma devono essere recepite dagli stessi, che hanno quindi margini di flessibilità nel modo di renderle operanti. I Regolamenti, invece, sono obbligatori in tutti i loro elementi e direttamente applicabili ai singoli soggetti in ciascuno degli Stati membri.)

Nella recezione della Direttiva gli Stati membri muovono dal comune riconoscimento che tutela della *privacy* e trattamento dei microdati per scopi scientifici sono entrambi ‘interessi meritevoli’, ma si distinguono nel ritenerli più o meno componibili. Sfortunatamente, l’Italia si colloca fra i paesi che propendono per una lettura antagonista; conseguentemente, detta una disciplina sull’accesso ai dati parecchio restrittiva.

La riflessione e le esperienze che seguono a livello europeo – fra Eurostat e Istituti Nazionali di Statistica (INS) –, estese all’OECD e all’UNECE e confortate dai risultati del progetto *Data Without Boundaries* (Jackson, 2019, e Trivellato, 2019), mostrano che il trattamento di microdati a fini scientifici è conciliabile con le garanzie di *privacy*. Viene così stabilito un quadro normativo nuovo. Da un lato sono adottati i Regolamenti 223/2009 (noto come *Statistical Law*) e 577/2013, che istituiscono il Sistema Statistico Europeo e definiscono classi di microdati, variamente informativi, e le condizioni alle quali vengono rilasciati per scopi scientifici. Dall’altro lato, è approvato il Regolamento 679/2016 (*General Data Protection Regulation*, GDPR), volto soprattutto a disciplinare gli obblighi di imprese e amministrazioni pubbliche e applicato dal maggio 2018. Le date non ingannino. Il GDPR, infatti, entra in vigore nel maggio 2016 e la sua applicazione è differita di due anni per consentire a imprese e amministrazioni pubbliche di adeguarvisi. D’altra parte, Regolamento 577/2013 e GDPR sono elaborati in parallelo – la prima bozza del GDPR è del gennaio 2012 – e in maniera coordinata.

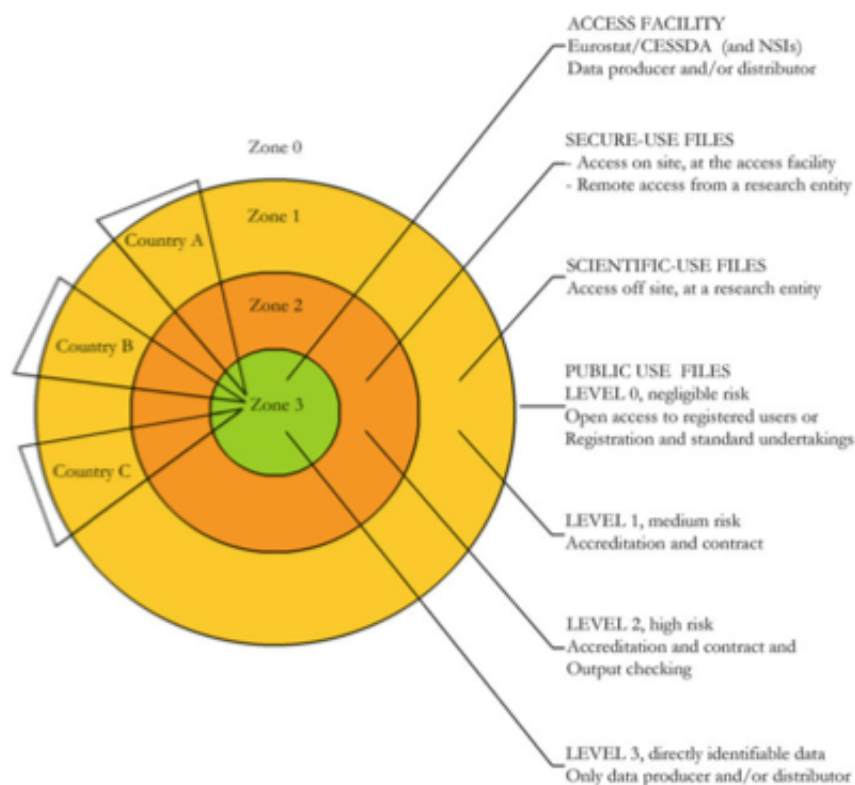
Si completa così il processo che, in mezzo secolo, porta da molteplici discipline nazionali a una disciplina comunitaria, da una normativa europea tramite direttive ad una tramite regolamenti, da rapporti fra INS e

comunità scientifica segnati dalla preoccupazione dei primi nei riguardi di potenziali intrusi a rapporti di *partnership*.

### Classi di microdati per la ricerca scientifica e condizioni di accesso.

La fig. 1 schematizza le tre classi di microdati disponibili per la ricerca scientifica e le condizioni alle quali sono accessibili ai ricercatori, con l'aggiunta di brevi note sulla corrispondenza fra i termini inglesi e quelli impiegati dall'Istat – tra parentesi – nonché sulle specifiche richieste dell'Istituto per il loro rilascio ([Direttiva Comstat](#) sulla quale tornerò).

Fig. 1: *Accesso a basi di micro-dati per scopi scientifici: classi di micro-dati, loro rischi di identificazione e requisiti per l'accesso*<sup>1</sup>



<sup>1</sup>Fonte: Trivellato (2019), p. 40. La figura fa riferimento al Sistema Statistico Europeo. Essa vale per un sistema statistico nazionale, una volta sostituiti a "Country A, Country B..." la sequenza dei file o archivi.

- A un livello di rischio di identificazione delle persone trascurabile si collocano i *Public Use Files* (File standard), che constano di “*record resi anonimi e predisposti in modo tale che le unità statistiche non possano essere identificate, direttamente o indirettamente, tenuto conto di tutti i pertinenti mezzi che possono essere ragionevolmente utilizzati da un terzo*”. Essi sono rilasciati previa registrazione e richiesta motivata del ricercatore.
- A un livello di rischio medio si colloca un primo insieme di “*confidential data for scientific purposes [ , that is] data which only allow for indirect identification of the statistical units*”, (dati elementari). Si tratta degli *scientific-use files* (file di microdati per la ricerca, MFR): dati elementari cui sono stati applicati metodi di controllo per la tutela della riservatezza che – considerate le finalità scientifiche per le quali sono prodotti – assicurano comunque il mantenimento di un elevato contenuto informativo. L'accesso a tali dati avviene presso l'entità di ricerca dei ricercatori interessati, previ accreditamento sia dell'entità di ricerca sia dei ricercatori e appropriati impegni di

riservatezza sottoscritti da entrambi i soggetti.

- A un livello di rischio elevato, infine, vi è il secondo insieme di dati elementari: i *secure-use files* (dati elementari cui non sono stati applicati metodi di controllo per la tutela della riservatezza). Palesemente, essi mantengono l'intera informazione di interesse. L'accesso dei ricercatori ai *secure-use files* avviene soltanto presso *Access facilities* (Laboratori), recandosi presso gli stessi o da remoto, da punti di accesso approntati presso le entità di ricerca interessate. In tal caso, le procedure di accreditamento si estendono all'adeguatezza della misure tecniche e organizzative per la gestione e la sicurezza dei dati. Inoltre il rilascio dei risultati delle analisi avviene soltanto a ricerca conclusa e previo controllo che non sono riferibili a singole persone.
- Al cuore dell'impianto stanno i Laboratori, presso i quali i file di dati sono prodotti o comunque mantenuti nella loro interezza, quindi identificabili anche direttamente. L'accesso è ristretto agli INS, agli altri enti produttori e agli enti distributori dei dati, se diversi dai precedenti. L'integrazione degli archivi è condotta a questo livello ed è decisiva per l'adeguatezza dell'offerta di basi di dati per scopi scientifici.

**L'accesso da remoto a *secure-use files* come modo ottimale per trattare dati elementari.**

Le pratiche di parecchi INS documentano che l'accesso da remoto a *secure-use files* è ormai diffuso (tab. 1). Esso è il preferito dai ricercatori – è ovvio, quando debbono trattare i file più informativi –, ma offre anche le massime garanzie di tutela della *privacy*. L'affermazione può forse sorprendere, ma soltanto l'accesso da remoto fornisce i cosiddetti '*Five safes*': (i) *entità di ricerca* ammissibili sicure; (ii) *ricercatori* ammissibili sicuri; (iii) *progetti di ricerca* ammessi sicuri; (iv) *dati* sicuri, perché non escono mai dai Laboratori; (v) *output* sicuri, perché rilasciati dopo aver accertato che non rivelano dati personali.

Tab. 1: Enti produttori/distributori di micro-dati con accesso da remoto in alcuni Paesi dell'UE<sup>1</sup>

Ambiti e Paesi	Ente	Caratteristiche salienti
<b>Ambiti settoriali</b>		
Germania	Research Data Centre (FDZ) della Federal Employment Agency, presso <u>Institute for Employment Research (IAB)</u>	Basi di dati in tema di lavoro
<b>Ambito generale: produttore e gestore l'INS</b>		
Danimarca	<u>Statistics Denmark</u>	Basi di dati largamente da registri
Olanda	<u>Statistics Netherlands (CBS)</u>	Basi di dati largamente da registri
<b>Ambito generale: gestore ad hoc di dati prodotti da INS e altri enti</b>		
Francia	<u>Centre d'Accès Sécurisé aux Données (CASD)</u>	<u>Basi di dati di INSEE; di Ministeri di Agricoltura, Finanze, Giustizia, Istruzione, Lavoro e Sanità; di Quêtelet-PROGEDO-Diffusion.</u>
Regno Unito	<u>UK Data Archive - Data Service Secure Lab</u>	Basi di dati di Office for National Statistics e UK Data <u>Archive</u>
Svezia	<u>Swedish Research Council - Registerforskning.se</u>	<u>Basi di dati di Statistics Sweden, Board of Wealth and Welfare, Karolinska Hospital e Swedeheart</u>

<sup>1</sup> Fonte: pertinenti siti web. Informazioni Aggiornate al 10.07.2019.

La Tab. 1 documenta anche una significativa novità. In tre paesi – Francia, Regno Unito e Svezia – l'accesso ai microdati da remoto avviene presso enti/centri ad hoc, ai quali confluiscono i dati degli INS e di altri enti.

---

## L'accesso a basi di microdati in Italia: incomprensioni, lentezze ... e una buona notizia.

Per un concorso di ragioni, l'Italia partecipa marginalmente alla riflessione e alle innovazioni della normativa dell'UE nel decennio 2007-'16. E il ritardo nell'adeguamento della normativa italiana al quadro europeo si fa pesante. Tra il Regolamento 577/2013 e la **Direttiva del Comstat del 2018 (che completa la disciplina** delineata nel decreto legislativo 97/2016, art. 5-ter) **passano più di 5 anni**, dei quali oltre i 3/4 chiamano in causa lentezze del legislatore e del Garante. La Direttiva del Comstat, poi, non è ancora operativa, sicché tra l'introduzione dell'accesso da remoto nei sei paesi elencati nella Tab. 1, prima del 2012 ([Gurke, Schiller e Gadouche](#)) e nel nostro paese corrono almeno 8 anni.

Per la disciplina dell'accesso ai microdati a fini di ricerca scientifica resta d'altra parte difficile comprendere perché sia stato – e continui ad essere – dato esorbitante rilievo al GDPR. Nelle disposizioni chiave al riguardo, infatti, esso non va oltre la fissazione di due principi generali: (i) il trattamento a tali fini di dati personali raccolti per altre legittime finalità è considerato compatibile con le finalità iniziali, (ii) a condizione che sia “*soggetto a garanzie adeguate per i diritti e le libertà dell'interessato [...tramite appropriate] misure tecniche e organizzative*” (artt. 5, par. 1, lett. b) e 89, par. 1). Paradossalmente, poi, a tale rilievo si è accompagnata opacità dell'adeguamento della normativa italiana al GDPR. Il legislatore italiano, infatti, mantiene l'impianto dell'ipertrofico Codice per la protezione dei dati personali e procede tramite revisioni – aggiunte, soppressioni, modifiche – di singoli articoli ([Codice](#)); questo, si noti, col GDPR che abroga tout court la Direttiva del 1995. Le conseguenze di questa scelta si colgono confrontandola con le modalità di adeguamento al DGPR adottate da altri paesi (ad es., il Regno Unito riscrive integralmente il pertinente capitolo del [Data Protection Act](#) in stretta aderenza al GDPR; La Francia rivede la [Loi Informatique et Libertés](#), rimarcando come per la sua comprensione occorra riferirsi sistematicamente al GDPR e, per un altro verso, dettando le disposizioni che si collocano entro i margini di flessibilità nazionali previsti dallo stesso). Non sorprendentemente, la nuova disciplina italiana della *privacy*, in particolare in tema di trattamento dei microdati per fini scientifici, è ancora ridondante, farraginosa, non esente da qualche incongruenza rispetto al GDPR.

La buona notizia, già anticipata, è che l'11 dicembre 2018 sono finalmente entrate in vigore le “**Linee guida per l'accesso a fini scientifici dei dati elementari del Sistan**”, adottate dal Comstat con la **Direttiva 11/2018**. Di massima esse sono allineate alle [Guidelines Eurostat](#), ma vi aggiungono confusi, talvolta gravosi appesantimenti.

**Al superamento di perduranti inadeguatezze occorre porre mano con sollecitudine e determinazione. Sul fronte della regolazione, l'opportunità è offerta dalla prossima revisione delle Regole deontologiche (prima chiamate Codici di deontologia e buona condotta) per trattamenti a fini statistici o di ricerca scientifica, che saranno adottate tramite la consultazione dei soggetti interessati – università e altre entità di ricerca, società scientifiche, Sistan –. Il comune impegno della comunità scientifica e del Sistan sarà decisivo per un cambio di passo.**

### La nuova sfida: alimentare la ricerca scientifica con basi di dati amministrativi.

La sfida che attende il Sistan e la comunità scientifica verte poi sulla piena valorizzazione dei dati di origine amministrativa, i cui molteplici vantaggi sono messi in luce da Paruolo in questo numero. Aggiungo due brevi notazioni: (i) le basi di dati amministrativi tornano particolarmente utili per la valutazione degli effetti di politiche: riguardando l'intera popolazione, forniscono informazioni su caratteristiche e risultati sia dei beneficiari della politica (trattati) sia di tutti i non beneficiari (non trattati), e consentono quindi di formare appropriati gruppi di confronto (controlli); (ii) l'utilizzo prevalente di dati amministrativi contraddistingue i sistemi statistici di parecchi paesi, tra i quali i *Nordic Countries* (vedi, ad es., [Statistics Denmark](#)) e l'orientamento verso un loro uso intensivo si sta affermando nell'intero Sistema Statistico Europeo ed è ribadito nel GDPR.

In Italia iniziative impegnative e innovative non mancano, dal *Programma di modernizzazione 2016* dell'Istat al recente programma *VisitInps*. Ma il quadro complessivo e la lentezza con la quale evolve restano preoccupanti. Un caso emblematico, che fortunatamente volge verso una conclusione positiva, è l'anagrafe nazionale della popolazione residente. Istituita con un decreto del 2005, è decollata soltanto nel 2016, con la costituzione di una struttura commissariale – il Team per la Trasformazione Digitale – che la completerà entro l'anno.

Tra gli ostacoli alla realizzazione di un sistema di archivi amministrativi (dove la parola chiave è 'sistema', che segnala l'essenziale requisito dell'integrazione) stanno il ritardo tecnologico delle amministrazioni pubbliche e l'assetto organizzativo inevitabilmente datato del Sistan. Un Sistan con l'Istat demiurgo; con novemila ipotetici Uffici di statistica, collocati in enti che vanno dai piccoli Comuni ai grandi ministeri e istituti produttori di dati; con il terreno per il coordinamento costituito soltanto dal Programma Statistico Nazionale.

Il tema è di grande latitudine. Concludo con sommari spunti per avviare la riflessione, incentrati su aspetti e obiettivi funzionali.

- La realizzazione di sistemi di archivi amministrativi si colloca ad appropriati livelli di centralizzazione e richiede la combinazione di competenze statistico-informatiche e sostanziali.
- I sistemi di archivi richiesti sono molto numerosi e tematicamente diversificati. È bene restare all'impianto tradizionale di produzione da parte di un solo ente, l'Istat, o conviene muovere verso un assetto contraddistinto da più enti, che li producono secondo standard comuni, e da un solo portale di accesso, come avviene in Francia, Regno Unito e Svezia?
- Alle amministrazioni ed enti che, per le funzioni che svolgono, necessitano di disporre di determinati archivi – o loro segmenti – va assicurato il periodico, tempestivo rilascio degli stessi.
- Per fini di ricerca scientifica, ai ricercatori ammissibili va assicurato l'accesso agevole e sicuro ad appropriate basi di dati tratte da sistemi di archivi.

Detto altrimenti, l'ineludibile centralizzazione del processo di produzione di archivi amministrativi deve essere affiancata da pratiche di tempestivo, amichevole e sicuro rilascio di (accesso a) microdati adeguati sotto il profilo sostanziale, identificati e aggiornati grazie a forum di produttori e utilizzatori dell'informazione. D'altra parte, più numerosi sono gli utilizzi degli archivi da parte di enti e ricercatori, maggiore è il vaglio della loro affidabilità. E più numerosi sono i *replication studies* di ricercatori indipendenti, più stringente è il processo che porta a spiegazioni credibili di processi sociali e politiche pubbliche.