

**Qualità dei dati amministrativi e dei registri statistici, strategie di
diffusione dei microdati**

Audizione dell'Istituto nazionale di statistica

**Commissione per la garanzia della qualità dell'informazione statistica
Roma, 23 settembre 2019**

Indice

1	Qualità dei dati amministrativi e dei registri statistici	5
1.1	I processi di gestione dei dati amministrativi	5
1.2	La qualità dei dati amministrativi	8
2	Miglioramenti nelle politiche di accesso ai microdati	12
2.1	Conservazione e accessibilità dei microdati	13
2.2	L'informazione sui microdati	14
2.3	Le Linee guida per l'accesso a fini scientifici ai dati elementari del Sistema statistico nazionale	14
2.4	Il rilascio dei microdati agli enti e uffici di statistica del Sistema statistico nazionale	16
3	RAF: un progetto strategico per la valorizzazione dei Registri	16
4	Tutela della privacy e impatto sulla statistica ufficiale	19

1 Qualità dei dati amministrativi

In questi ultimi anni l'Istat ha profondamente innovato i processi di produzione con lo scopo di ampliare l'offerta informativa e aumentare l'efficienza del sistema, come indicato dalle raccomandazioni degli organismi comunitari e internazionali. Il progressivo utilizzo di microdati da fonte amministrativa e la costruzione dei registri statistici, che permettono una descrizione più dettagliata dei fenomeni in una prospettiva multi-dominio, hanno reso necessario lo sviluppo di nuove formalizzazioni e nuove metodologie.

Le innovazioni, inizialmente sperimentali, si avviano verso una progressiva fase di standardizzazione sia dal punto di vista gestionale che metodologico. La necessaria documentazione della qualità nei suoi molteplici aspetti si va, quindi, consolidando.

1.1 I processi di gestione dei dati amministrativi¹

Nel Programma statistico nazionale del triennio 2017-2019 - Aggiornamento 2018-2019² sono 137 i lavori dell'Istat che dichiarano l'utilizzo di dati da fonte amministrativa su un totale di 319, ovvero il 43%. Una percentuale piuttosto elevata considerando tale misura come grado di dipendenza della statistica ufficiale da dati prodotti esternamente all'Istituto. Lo scenario presenta, quindi, una cresciuta collaborazione con gli Enti titolari di dati amministrativi, per la maggior parte afferenti alla Pubblica Amministrazione, ed evidenzia la necessità di mantenere una coesione di intenti che garantisca la continuità della produzione anche nel rispetto del contesto internazionale.

Per fornire un'ulteriore misura quantitativa della mole di microdati da fonte amministrativa che l'Istat acquisisce a scopo statistico, si riporta nella seguente tabella l'andamento degli ultimi tre anni. Il numero di Enti titolari va progressivamente aumentando: nel 2019 sono 55 tra Ministeri, Enti, Agenzie che collaborano stabilmente con l'Istat per coordinare il passaggio dei dati, 194 le fonti amministrative da cui le 457 forniture vengono periodicamente estratte e trasmesse all'Istat.

I valori tendono nel tempo a stabilizzarsi, dopo il periodo di forte crescita avviato circa 20 anni fa e sollecitato dalla fase esplorativa delle nuove fonti utilizzabili a scopi statistici

¹ Si riportano gli aggiornamenti rispetto alle Relazioni precedenti.

² Programma Statistico Nazionale attualmente in vigore in base al DECRETO DEL PRESIDENTE DELLA REPUBBLICA 20 maggio 2019 Approvazione del Programma statistico nazionale e degli altri atti di programmazione della statistica ufficiale 2017-2019 - Aggiornamento 2018-2019 (pubblicato sul Supplemento Ordinario n. 30 alla Gazzetta Ufficiale - serie generale - n. 165 del 16 luglio 2019).

e la sempre più proficua collaborazione degli Enti titolari dei dati amministrativi nell'obiettivo istituzionale condiviso di fornire informazione di qualità al Paese.

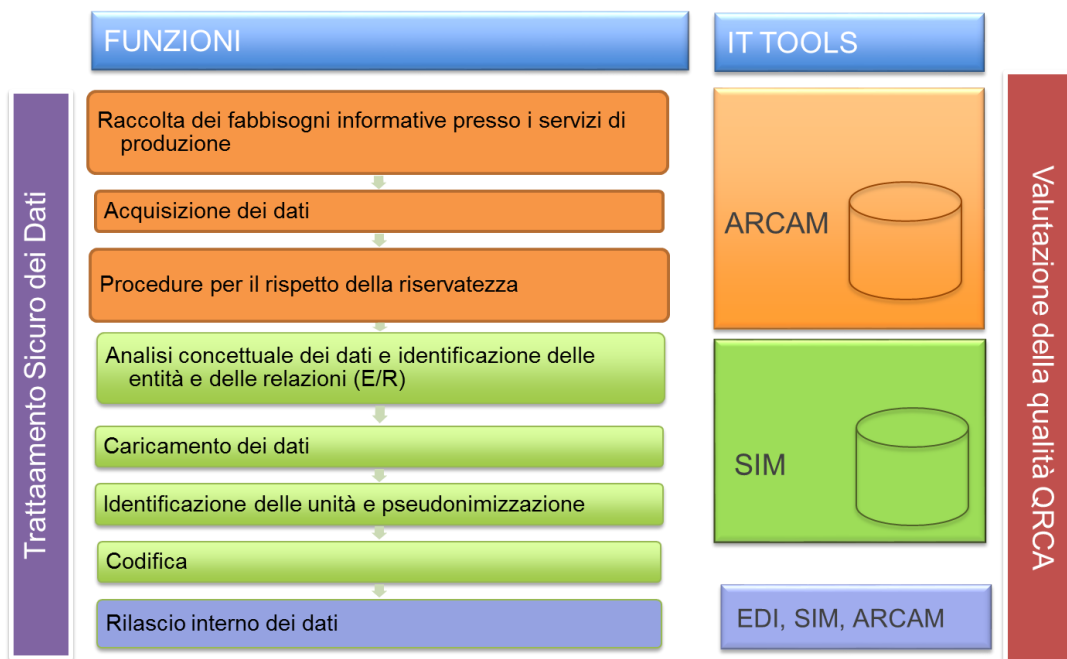
Tabella 1. Numerosità degli Enti titolari, delle fonti e delle forniture di dati amministrativi in Istat

Anno	Enti titolari	Fonti amministrative	Forniture periodiche
2017	47	182	452
2018	53	192	476
2019	55	194	457

Il sistema di gestione di questa ampia mole di dati, per la maggior parte dati personali, ha reso necessaria la costruzione e l'adozione di specifici IT TOOL che garantiscono l'integrità dei dati e il rispetto della normativa sulla riservatezza - *by design e by default* - in un'ottica sistemica.

Le funzioni che la struttura centralizzata preposta alla gestione dei dati amministrativi svolge sono sintetizzate in Figura 1. Il processo prende avvio dalla funzione di raccordo tra le esigenze dei processi di produzione statistica dell'Istat e la disponibilità dei dati amministrativi, a questo proposito annualmente viene effettuata una rilevazione delle esigenze informative presso i settori di produzione. Tali esigenze vengono poi conciliate anche secondo il principio di "minimizzazione dei dati" (dati personali adeguati, pertinenti e limitati – art. 5 Regolamento UE 2016/679) e, in accordo con gli Enti fornitori, si programmano i flussi di dati dalle varie fonti. In particolare si definiscono: i contenuti, la periodicità di invio dei dati (mensili, trimestrali, annuali, ...), la tipologia di invio (dati provvisori e definitivi o invii in un'unica soluzione) e la tempistica di acquisizione in funzione delle esigenze di tempestività della produzione delle statistiche. Una volta definito il calendario si procede con la funzione di acquisizione e di monitoraggio. Le attività sono supportate, in questa fase, dal sistema denominato ARCAM costituito da un portale web a disposizione degli Enti fornitori di dati che, utilizzando delle specifiche credenziali possono accedere e inviare i dataset programmati. Un secondo applicativo ARCAM, interno all'Istituto e che accede ai soli metadati descrittivi delle forniture, supporta il lavoro di back office di gestione e monitoraggio dell'acquisizione delle forniture di dati.

Figura 1. Il processo di gestione dei dati amministrativi utilizzati a fini statistici in Istat



Il *repository* centralizzato all'interno della rete intranet provvede alla conservazione dei dati.

I dati acquisiti seguono poi due strade: 1) i macrodati o dataset non contenenti dati personali vengono rilasciati in forma grezza (IT TOOL EDI) ai processi di produzione che ne hanno fatto richiesta; 2) i microdati vengono specificamente trattati nell'ambito del SIM (Sistema Integrato dei Microdati) che ne garantisce la riservatezza. Nello specifico si effettua, in prima acquisizione, un'analisi concettuale del dataset; i dati vengono caricati e gli identificativi diretti vengono separati dagli attributi e immagazzinati in distinti "schemi" e tabelle del database Oracle; si applica il processo di pseudonimizzazione dei dati; i metadati vengono trattati e mantenuti. Completato il processo in SIM, si rendono disponibili i dati ai processi di produzione di titolarità dell'Istat che ne hanno fatto richiesta (dichiarandoli nel Piano Statistico Nazionale) e ai ricercatori Istat autorizzati esplicitamente al trattamento dei dati personali per le finalità statistiche.

Due funzioni di tipo trasversale presidiano il processo: la verifica della sicurezza e la documentazione della qualità, di quest'ultima si tratterà nel seguente paragrafo.

1.2 La qualità dei dati amministrativi

Il ciclo di vita dei dati amministrativi utilizzati a scopo statistico (Figura 2) attraversa tre Macro processi³: 1) presso l'Ente titolare dei dati amministrativi dove, per l'attuazione di una normativa o di un processo amministrativo, si generano i dati; 2) presso l'Istat – Settore centralizzato di acquisizione dei dati amministrativi; 3) presso i settori di produzione dell'Istat, suddivisi per domini, che producono le statistiche di output.

Figura 2. Il ciclo di vita dei dati amministrativi per la produzione delle statistiche



L'Istat presidia la qualità dei 2 Macro processi interni all'Istituto ma svolge un ruolo attivo anche nell'ambito della prima fase: la collaborazione con gli Enti per la condivisione delle finalità statistiche, attuata anche mediante la sottoscrizione di accordi, sollecita la riduzione della distanza tra dati amministrativi e dati statistici, quando possibile, e la conseguente minimizzazione degli errori sia rispetto alle unità che alle variabili⁴.

Il macro processo che conclude il ciclo di vita, in cui eventualmente i dati amministrativi vengono combinati tra loro o con dati da indagine, garantisce la validazione e la

³ Istat (2016) Linee guida per la qualità dei processi statistici che utilizzano dati amministrativi Versione 1.1 (<https://www.istat.it/files/2010/09/Linee-Guida-fonte-amministrativa-v1.1.pdf>)

⁴ Zhang L.C. (2012). Topics of statistical theory for register-based statistics and data integration. Statistica Neerlandica, Vol 66, nr.1, pp. 41-63.

produzione ufficiale delle statistiche che possono essere diffuse all'esterno in forma di microdati o macrodati nelle modalità previste⁵.

Per il Macro processo 2, è stato sviluppato in Istat un sistema di documentazione che permette di focalizzare i concetti di riferimento della "qualità dell'input" per i processi di produzione delle statistiche che utilizzano dati amministrativi⁶.

Il sistema, entrato a regime a novembre del 2018, denominato *Quality Report Card dei dati Amministrativi* - QRCA - è disponibile per gli utenti interni Istat all'indirizzo <http://qrca.istat.it>

Gli obiettivi della QRCA sono:

1. Valutare la qualità delle forniture dei dati amministrativi per: a) supportare il processo di acquisizione; b) misurare l'usabilità statistica dei dati amministrativi acquisiti; c) monitorare le variazioni inattese della qualità/usabilità dei dati acquisiti.
2. Controllare la coerenza del sistema di integrazione.
3. Produrre il sistema di documentazione della qualità dei dati amministrativi e del processo di integrazione.

Il portale mette a disposizione tutti i metadati descrittivi delle forniture programmate per ciascun anno (2017, 2018, 2019) e lo stato del monitoraggio dell'acquisizione di ciascuna. Sono disponibili, inoltre, i contenuti dei dataset in termini di tipi di unità, di variabili, di classificazioni amministrative al fine di valutarne l'usabilità. Oltre alla documentazione di base, sono proposti degli indicatori di qualità.

Il Framework della qualità adottato in Istat in cui l'informazione è organizzata è brevemente descritto in Tabella 2 e trae origine da un lavoro proposto da Statistics Netherlands e sviluppato nell'ambito del progetto internazionale BlueEts⁷.

Le informazioni pubblicate nella prima versione prototipale, per ciascuna Iperdimensione della qualità, sono le seguenti:

⁵ Per le modalità di accesso ai microdati si veda la precedente Audizione dell'Istat presso la Commissione per la garanzia della qualità dell'informazione statistica Conservazione e accessibilità dei microdati prodotti dall'Istat e cenni all'ambito del Sistan, Roma, 22 settembre 2017.

⁶ Cerroni, F., Di Bella, G., & Galiè, L. (2014). Evaluating administrative data quality as input of the statistical production process, *Rivista di statistica ufficiale* 1-2/2014, 117-146 https://www.istat.it/it/files/2014/10/Articolo-7_Evaluating-administrative....pdf.

⁷ Daas et al. (2009) Checklist for the Quality evaluation of AD Sources. Discussion paper 09042, Statistics Netherlands. Daas et al. (2011) Reports on methods preferred for the quality indicators of administrative data sources, Deliverable 4.2 of Workpackage 4 of the BLUE-ETS project. CBS, Netherlands, SSB, Norway, Istat, Italy, SCB, Sweden.

FONTI

- Identificazione della Fonte, dell'Ente titolare
- Trattamento previsto per i dati
- Serie storica disponibile
- Rilevanza della fonte in termini di estensione di uso in Istat
 - Elenco dei lavori PSN che utilizzano la fonte
 - Normativa UE il cui adempimento dipende dalla fonte

METADATI

- Lista delle variabili amministrative
- Classificazioni amministrative presenti nei dataset forniti (variabili categoriche)
- Tipi di unità presenti (Individui, Unità economiche, Luoghi)

DATI

- Identificazione delle forniture, periodicità, riferimenti temporali
- Monitoraggio dell'acquisizione e del trattamento previsto sui dati al fine di verificare la disponibilità per l'accesso da parte degli utenti autorizzati
- *Technical checks* per la verifica della conformità dei dati ricevuti rispetto a quelli richiesti in termini di variabili acquisite (completezza della fornitura)
- Monitoraggio della qualità: confronti in serie storica del numero di record di ciascun file che compone la fornitura come indicatore di *warning*
- Misure della qualità delle variabili di linkage (variabili identificative disponibili nella fonte e percentuale di valori mancanti)
- Misure del monitoraggio della qualità delle procedure di record linkage (misure deterministiche)
- Puntualità dell'ente fornitore (rispetto della tempistica di acquisizione concordata)
- Percentuale di valori mancanti per le variabili (anche in serie storica)
- Controlli di completezza delle decodifiche delle classificazioni
- Tempestività dei dati in riferimento all'Ente (intervallo temporale tra la data dell'ultimo evento registrato nella fornitura e la data di acquisizione)
- Tempestività dei dati in riferimento alla disponibilità finale (intervallo temporale tra la data dell'ultimo evento registrato nella fornitura e la data di disponibilità dei dati per gli utenti interni)

RAPPORTI RIASSUNTIVI

- Tabella complessiva del monitoraggio dell'acquisizione e del trattamento di tutte le fonti amministrative acquisite
- Tabelle di riepilogo del Processo di acquisizione tramite ARCAM: Enti titolari, Archivi e Forniture per anno di programmazione delle acquisizioni e per Ente fornitore
- Tabelle di riepilogo del Processo di trattamento in SIM: Archivi e Forniture per anno di programmazione delle acquisizioni e per tipo di trattamento in SIM; Dettaglio degli archivi per tipo di trattamento in SIM

Il portale di documentazione è aggiornato in tempo reale grazie alla strategia di implementazione adottata: lo strumento BI standard dell'Istat, Microstrategy, accede ai DB di gestione dei dati amministrativi e sfrutta i metadati esistenti. Un investimento iniziale di analisi dei metadati disponibili, di standardizzazione e di riuso per le finalità non previste, la costruzione del sistema di produzione dei report, hanno permesso di realizzare in soli due anni la QRCA con risultati superiori a quelli previsti. I concetti statistici connessi all'uso dei dati amministrativi sono ora accessibili a tutti i ricercatori Istat. Inoltre, in aggiunta alle funzioni inizialmente previste, è risultata di grande utilità la funzione di verifica dello stato di disponibilità delle forniture che ha ridotto drasticamente la necessità di comunicazioni ad hoc.

Tabella 2 - Framework della qualità dei dati amministrativi utilizzati a scopo statistico

IPERDIMENSIONE	DIMENSIONE
FONTE	Informazioni sull'archivio
Informazioni necessarie a gestire il processo di acquisizione dei dati con lo scopo di valutare e migliorare la qualità dei dati acquisiti	Rilevanza della fonte (utilizzi statistici)
	Questioni connesse alle norme di uso (accessibilità)
	Relazioni e feedback con i titolari delle fonti
METADATI	Chiarezza e interpretabilità dei dati
Informazioni per la valutazione della qualità a livello concettuale e di processo presso il titolare della fonte	Comparabilità tra concetti amministrativi e concetti statistici (unità e variabili)
	Descrizione del processo di acquisizione e trattamento (da parte del titolare)
DATI	Accordi per l'acquisizione dei dati, monitoraggio dell'acquisizione e del trattamento
Valutazione della qualità dei dati acquisiti	Technical checks
	Integrabilità/Integrazione
	Accuratezza e coerenza interna
	Completezza (copertura per le unità, <i>missing values</i> per le variabili)
	Aspetti temporali

Figura 3. Homepage della QRCA (Quality Report Card dei dati Amministrativi)

Quality Report Card dei dati Amministrativi

QRCA | [Home](#) | [Selezione Archivio](#) | [Fonte](#) | [Metadati](#) | [Dati](#) | [Rapporti riassuntivi](#) | [Gestione Portale](#)

Sistema di documentazione dei dati amministrativi acquisiti in Istat

La Quality Report Card dei dati Amministrativi (QRCA) è un Sistema di documentazione dei dati amministrativi, acquisiti dall'Istat da fonti esterne.

Lo scopo è di supportare i processi di produzione delle statistiche e potenziare, laddove possibile, l'uso di questi dati.

Il Sistema si alimenta in modo automatico dai DB di gestione dei dati amministrativi in Istat: attualmente **ARCAM** (il DB del portale di acquisizione dei dati amministrativi) e **SIM** (il Sistema di integrazione dei dati amministrativi).

L'informazione è organizzata secondo il framework della qualità sotto riportato. Le dimensioni attualmente implementate (**Versione prototipale 1.0**) sono evidenziate in grassetto.

Dopo l'accreditamento, **selezionare l'archivio di interesse** e si accederà alle informazioni relative alla **Fonte** amministrativa (trattamento e Rilevanza ovvero lavori PSN che utilizzano l'archivio). Si potrà poi navigare nei report dei **Metadati** (informazioni sui contenuti dell'archivio) e dei **Dati** (stato dell'acquisizione delle forniture, puntualità e tempestività, controlli di qualità dei dati, indicatori sul processo di integrazione).

Ulteriori informazioni sono disponibili nei footer. Per informazioni, suggerimenti, malfunzionamenti scrivere a qrca@istat.it.

SI CONSIGLIA DI NON UTILIZZARE IL BROWSER INTERNET EXPLORER

Framework della qualità dei dati amministrativi acquisiti in Istat

IPERDIMENSIONE	DIMENSIONE
FONTE <i>Informazioni necessarie a gestire il processo di acquisizione dei dati con lo scopo di valutare e migliorare la qualità dei dati acquisiti. Usi statistici</i>	Informazioni sull'archivio
	Rilevanza della fonte (utilizzi statistici - PSN e Regolamenti)
	Questioni connesse alle norme di uso (accessibilità)
	Accordi specifici con l'ente
	Relazioni e feedback con i titolari delle fonti
METADATI <i>Informazioni per la valutazione della qualità a livello concettuale e di processo presso il titolare della fonte</i>	Chiarezza e interpretabilità dei dati
	Comparabilità tra concetti amministrativi e concetti statistici (unità e variabili)
	Descrizione del processo di acquisizione e trattamento (da parte del titolare)
DATI <i>Valutazione dei dati acquisiti, della loro qualità e del processo di trattamento</i>	Caratteristiche e stato dell'acquisizione delle forniture
	Aspetti temporali
	Controlli tecnici
	Integrabilità/Integrazione
	Accuratezza e coerenza interna
	Completezza delle variabili e delle unità

[Mappa delle informazioni](#) | [Note tecniche](#) | [Informazioni per la navigazione](#) | [Glossario](#) | [Bibliografia](#) | [Team](#) |

In particolare la componente DB metadati di ARCAM fornisce le informazioni sulle identificazioni e descrizioni dei data set la cui acquisizione è programmata per l'anno, lo stato ed i tempi di acquisizione; il DB SIM nella componente metadati (SISME) fornisce le informazioni sui trattamenti dei dati, sulla descrizione dei dati in termini di tipi di unità e di variabili, sullo stato di esecuzione dei trattamenti di pseudonimizzazione. Ulteriori informazioni derivano da specifiche tabelle di raccordo mantenute periodicamente e da alcuni macrodati le cui procedure di calcolo vengono eseguite periodicamente a valle dei caricamenti, infine vengono utilizzate alcune tabelle disponibili automaticamente nel Dizionario di Oracle.

Per quanto riguarda le relazioni con gli altri sistemi disponibili in Istat che possono contribuire ad arricchire l'informazione e rendere sempre più coerente il sistema, molti passi sono stati fatti. Il collegamento con il DB del Programma Statistico Nazionale, che fornisce l'abbinamento tra lavoro statistico e fonte amministrativa, è attivo ma ancora in fase di automatizzazione. E' in corso un processo di armonizzazione dei metadati tra la QRCA e il Sistema Informativo sulla Qualità dei processi di produzione Istat (SIQual, disponibile per gli utenti all'indirizzo <http://siqua.istat.it>): lo scopo è condividere e non duplicare l'informazione comune e di connessione dei due sistemi. Inoltre, è stata avviata la collaborazione con il Sistema Unitario dei Metadati (SUM) che gestisce i metadati dell'Istat: in questo caso il primo obiettivo è di rendere disponibili nella QRCA le funzioni di trasformazione delle classificazioni amministrative in classificazioni statistiche, laddove disponibili.

2 Miglioramenti nelle politiche di accesso ai microdati

La crescente disponibilità di basi di microdati risultanti da indagini e fonti amministrative rappresenta una componente fondamentale del patrimonio informativo dell'Istituto.

Il forte incremento della domanda di microdati, da parte delle istituzioni e della comunità scientifica in generale per analisi a carattere tematico (economico, sociale, ambientale ecc.) e territoriale, ha portato l'Istat a rivedere le logiche di conservazione dei dati elementari d'indagine e a valutare nuove politiche di diffusione dei microdati, anche attraverso l'adozione di strumenti in grado di tutelare la *privacy* dei soggetti interessati.

2.1 Conservazione e accessibilità dei microdati

L'Istat, cogliendo le indicazioni provenienti dall'Unione europea, si è impegnato assiduamente, nel rispetto delle norme vigenti⁸, ad ampliare quanto più possibile l'accesso ai microdati d'indagine di cui è titolare.

I microdati validati prodotti sono archiviati centralmente nell'ARchivio dei MIcroDATi (ARMIDA), con il duplice scopo di conservare i microdati e i relativi metadati delle rilevazioni condotte dall'Istituto e favorire il riutilizzo degli stessi per finalità statistiche e di ricerca da parte di utenti sia interni sia esterni all'Istat. Alla fine del 2018 l'archivio ARMIDA raccoglie e documenta i microdati validati di 259 indagini (di cui 104 cessate o sospese), per un totale di 14.876 file di microdati.

Nel corso dell'anno l'Istat ha avviato un progetto per la reingegnerizzazione di ARMIDA avente principalmente il compito di definire nuovi requisiti utente, funzionalità e architettura dei contenuti, anche con l'obiettivo di consentire ai ricercatori interni una modalità di accesso semplificato finalizzata all'analisi statistico-economica, statistico-sociale e demografica. Da questo punto di vista si aprono per ARMIDA nuovi scenari di utilizzo e di maggiore integrazione con altri sistemi informativi e registri interni all'Istituto al fine di consentire il riuso dei dati e relativi metadati nelle diverse fasi del processo produttivo.

Come già ricordato nelle precedenti audizioni, accanto alla tradizionale offerta di File standard, già da alcuni anni l'Istat diffonde i File per la ricerca (MFR)⁹, realizzati specificamente per soddisfare esigenze conoscitive da parte dei ricercatori di Università o Enti di ricerca e, dal 2013, ha avviato la produzione di File a uso pubblico (PUF, in Istat chiamati *micro.Stat*), collezioni di dati elementari scaricabili liberamente e gratuitamente dal sito istituzionale. Alla produzione di questi file si associa inoltre la possibilità per ricercatori e studiosi di condurre in autonomia analisi statistiche sui microdati validati, grazie all'accesso al Laboratorio ADELE.

Nel 2018 le richieste di rilascio di microdati autorizzate dall'Istat sono state 694, per un totale di 2.555 file rilasciati. In particolare, i File per la ricerca (MFR) autorizzati sono stati 162 e 57 risultano i progetti attivi presso il Laboratorio ADELE (in lieve aumento rispetto ai 51 del 2017). Scende del 43,7% il numero delle richieste di File standard (che si attestano a 143 nel 2018) in quanto non più prodotti dall'Istituto e progressivamente sostituiti dai file MFR, mentre aumenta quello dei File ad uso pubblico scaricati direttamente dal sito dell'Istat (pari a 9.169 nel 2018), che quasi si triplica rispetto

⁸ Art. 10, comma 2, del d.lgs. 322/1989 e Art. 2-ter Nuovo Codice Privacy – d.lgs. 196/2003, integrato con le modifiche introdotte dal d.lgs. 101/2018 in materia di trattamento dei dati personali.

⁹ L'Istat li diffonde dal 2010 in conformità all'Art. 6 del Regolamento (CE) n. 831/2002, che ha introdotto l'accesso a "microdati anonimi" ottenuti da alcune indagini, poi abrogato dal successivo Regolamento (UE) n. 557/2013 che ha definito l'accesso ai dati riservati destinati a fini scientifici.

all'anno precedente, segno dell'interesse sempre maggiore, anche da parte di un pubblico più "generalista", verso i microdati di rilevazioni quali, ad esempio, Forze di lavoro, Uso dell'ICT nelle imprese, Aspetti della vita quotidiana e Spese delle famiglie.

2.2 L'informazione sui microdati

I file di microdati sono collezioni di dati elementari relativi alle rilevazioni. L'Istat li rilascia a titolo gratuito per finalità di studio e ricerca e per finalità statistica scientifica, nel rispetto di quanto consentito dalla normativa vigente (art. 19 reg. (CE) n. 223/2009; art. 6, comma 1, lett. b) e art. 10, comma 2, d.lgs. n. 322/1989; art. 5-ter del decreto legislativo n. 33/2013; art. 7, comma 1, "Regole deontologiche per trattamenti a fini statistici o di ricerca scientifica effettuati nell'ambito del Sistema statistico nazionale").

L'Istat rilascia diverse tipologie di microdati. Un'apposita sezione del sito istituzionale guida i ricercatori tra le informazioni sui file di microdati resi disponibili, dando la possibilità di visionare la gamma di dati, completa di tutte le metainformazioni relative a variabili, tracciati, modalità e moduli di accesso, normativa di riferimento.

2.3 Le Linee guida per l'accesso a fini scientifici ai dati elementari del Sistema statistico nazionale

La Direttiva n. 11 del 7 novembre 2018 del Comitato di indirizzo e coordinamento dell'informazione statistica (Comstat) adotta le "Linee guida per l'accesso a fini scientifici ai dati elementari del Sistema statistico nazionale", in attuazione dell'art. 5-ter del decreto legislativo n. 33 del 14 marzo 2013, introdotto dal decreto legislativo 25 maggio 2016, n. 97. Le Linee guida fissano le condizioni in base alle quali gli enti e gli uffici del Sistan possono consentire ai ricercatori l'accesso per fini scientifici ai dati elementari, di cui sono titolari, privi di riferimenti che permettano l'identificazione diretta delle unità statistiche.

La procedura definita dal D.lgs. prevede in sintesi che:

- l'ente di appartenenza dei ricercatori sia riconosciuto come ente di ricerca dal Comstat sulla base di criteri prestabili (art. 5-ter comma 1, lett. a) del D.lgs.) o faccia parte dell'elenco degli Enti di ricerca riconosciuti da Eurostat (Regolamento (UE) n. 557/2013).
- i ricercatori dell'ente di ricerca riconosciuto presentino una Proposta di ricerca per il cui svolgimento è necessario ricorrere a dati elementari e sottoscrivano una Dichiarazione individuale di riservatezza. La Proposta di ricerca è valutata dall'Ente del Sistan titolare dei dati sulla base dei criteri stabiliti dal Comstat (art. 5-ter comma 1, lett. c) del D.lgs.).
- la Proposta di ricerca approvata consenta ai ricercatori indicati nella stessa Proposta di accedere (art. 5-ter comma 2 del D.lgs.):
 - a) ai dati elementari cui sono stati applicati metodi di controllo per la tutela della riservatezza (file per la ricerca, di seguito file MFR);

- b) ai dati elementari cui non sono stati applicati metodi di controllo per la tutela della riservatezza, nell'ambito di appositi Laboratori, accessibili anche da remoto, per l'analisi di dati elementari, diversi dal laboratorio ADELE che si trova nella sede centrale dell'Istat e da quelli dislocati negli uffici delle sedi territoriali Istat.

Il riconoscimento di un'organizzazione quale Ente di ricerca è condizione necessaria affinché i ricercatori dell'Ente possano richiedere l'accesso ai file di dati elementari, resi noti dall'Ente del Sistan sul proprio sito istituzionale, tramite Laboratorio o file MFR, e che l'Ente possa sottoscrivere un Protocollo di ricerca con l'Ente del Sistan titolare dei dati per condurre un progetto congiunto che preveda l'uso di dati elementari.

La procedura di riconoscimento, prevede che l'Ente del Sistan titolare dei dati che riceve la domanda, ne valuta la conformità secondo i criteri previsti nelle linee guida, e invii poi al Comstat l'esito dell'istruttoria per l'approvazione. L'Istat ha ricevuto 16 richieste di riconoscimento, due delle quali approvate nell'ultima seduta del Comstat del 02/04/2019; le restanti 14 candidature, sottoposte già a valutazione, sono in attesa della nomina del nuovo Comstat per poter essere approvate.

Le organizzazioni che possono essere prese in considerazione per il riconoscimento (art. 5-ter, comma 1, lett. a), del D.lgs.) sono università o loro strutture interne di ricerca, enti di ricerca o loro strutture interne di ricerca, istituzioni pubbliche o private o loro strutture interne di ricerca.

Al fine di dare attuazione al D.lgs l'Istat ha avviato una revisione complessiva del percorso autorizzatorio per l'accesso ai dati e una revisione della modulistica, tanto per la richiesta dei file MFR quanto per l'accesso al Laboratorio ADELE.

Il 27 febbraio 2019 è stata costituita una Task force avente il compito di individuare le attività necessarie a dare completa e tempestiva attuazione alle "Linee guida per l'accesso a fini scientifici ai dati elementari del Sistema statistico nazionale (Sistan)". In particolare, la Task force è stata incaricata di formulare proposte operative con particolare riferimento all'attuazione da parte dell'Istat e dei soggetti del Sistan delle disposizioni dettate dalle predette "Linee guida" in materia di riconoscimento degli Enti di ricerca, di valutazione delle proposte di ricerca, di costituzione e funzionamento dei laboratori per l'accesso ai dati.

In sintesi, la Task force ha provveduto a:

- i. individuare i requisiti necessari per rispondere alle esigenze di compilazione assistita della modulistica, aggancio alle indagini richieste in Armida, invio automatico della richiesta al fine di attivare l'iter autorizzativo
- ii. apportare una revisione normativa a tutta la documentazione interessata (moduli di accesso, Linee Guida al Laboratorio ADELE, pagine del sito istituzionale)

- iii. definire la procedura interna per il riconoscimento degli enti di ricerca e pubblicazione degli esiti delle valutazioni positive di riconoscimento sul sito istituzionale
- iv. ragionare sui requisiti necessari per la costituzione di laboratori per l'accesso remoto ai dati elementari e sull'opportunità di procedere ad una successiva e necessaria sperimentazione al fine di produrre una documentazione di dettaglio.

2.4 Il rilascio dei microdati agli enti e uffici di statistica del Sistema statistico nazionale

Le modifiche del quadro normativo in materia di privacy, delle disposizioni in materia di liceità del trattamento dei dati personali e delle relative disposizioni che ne regolamentano lo scambio, disciplinato dall'art. 6 del decreto legislativo n. 322/1989, dall'articolo 8 del Codice di deontologia e dalla Direttiva n. 9 del Comstat, hanno richiesto una riflessione relativa anche all'interscambio di dati all'interno del Sistema statistico nazionale.

Con la pubblicazione in Gazzetta Ufficiale delle Regole deontologiche per i trattamenti a fini statistici o di ricerca scientifica effettuati nell'ambito del Sistema statistico nazionale, l'Autorità ha abrogato l'articolo 8 del Codice, in materia di comunicazione dei dati tra soggetti del Sistan, poiché non conforme al rinnovato quadro normativo, in quanto ai sensi dell'articolo 6, par. 3 del Regolamento e dell'art. 2-ter del d.lgs n. 196/2003, "si richiede una norma di legge o, nei casi previsti dalla legge, di regolamento per la comunicazione di dati tra soggetti pubblici o tra soggetti che svolgono compiti di interesse pubblico, quali sono i soggetti del Sistan".

A tal fine è stato necessario l'adeguamento della scheda allegata alla direttiva Comstat n.9/2004. L'intervento ha riguardato sostanzialmente la revisione della terminologia delle sezioni 2 e 3 e l'introduzione della sezione 4 - base giuridica del trattamento dei dati personali rientranti in particolari categorie (ex dati sensibili) e di dati relativi a condanne penali e reati (ex dati giudiziari) indicando, a seconda della tipologia dell'informazione, la disposizione di legge o di regolamento che legittima il richiedente al trattamento dei dati. L'approvazione della modifica, alla luce delle evoluzioni normative in materia di protezione dei dati personali e della riformulazione dell'art. 6 commi 1, lett. b e 4 del d.lgs. n.322/89, è avvenuta in sede di Comitato (nella seduta del 2 aprile 2019).

3 RAF: un progetto strategico per la valorizzazione dei Registri

L'Istat ha recentemente dato avvio alle attività di un progetto per la realizzazione del prototipo della nuova infrastruttura informativa Register based Analytics Framework (RAF), volto alla valorizzazione del potenziale informativo del sistema dei registri statistici dell'Istat.

L'esigenza di una nuova piattaforma statistica e tecnologica di diffusione nasce dalla peculiarità della nuove basi dati statistiche, svincolate dal disegno d'indagine, fondate su dati elementari di intere popolazioni di riferimento e poste a sistema mediante integrazione concettuale e statistica delle unità che le compongono.

Su un piano generale, il progetto rappresenta una componente rilevante dell'evoluzione, anche in senso organizzativo, del Sistema Integrato dei Registri (SIR) come snodo della nuova strategia Istat di produzione adottata formalmente e in modo strutturato nel 2016 ma già ampiamente presente, seppure in modo non sistematico, in alcuni domini statistici.

Più in dettaglio, il progetto è fortemente caratterizzante il più ampio obiettivo strategico 2020-2022 di valorizzazione del potenziale informativo dei registri statistici e del sistema derivante dalla loro integrazione (SIR), presente all'interno del PG1 relativo al prossimo triennio.

L'obiettivo a breve termine è la progettazione e lo sviluppo del prototipo di una nuova infrastruttura statistica e tecnologica a supporto di un adeguato sfruttamento del potenziale informativo del sistema dei registri statistici dell'Istat.

La nuova infrastruttura informativa, denominata *Register based Analytics Framework* (RAF), dovrà essere progettata e realizzata raccogliendo i requisiti e le effettive esigenze delle diverse tipologie di utenti e finalizzata alla migliore fruibilità pubblica della nuova offerta di informazione statistica basata sui registri e sulla loro integrazione.

Il nuovo modello di produzione di dati genera infatti un'offerta potenziale d'informazione statistica di grande estensione e profondità, particolarmente rilevante in considerazione della crescente domanda di misurazioni di fenomeni complessi a fini di statistiche ufficiali, di ricerca e a supporto dell'orientamento e della valutazione (ex-ante ed ex-post) delle policy pubbliche.

In tale contesto, l'obiettivo specifico di progettazione e sviluppo del prototipo del RAF è perseguito attraverso l'organizzazione di un complesso sistema di attività, dedicate ad affrontare i diversi aspetti, statistico-metodologici, tecnologici e informatici, di diffusione e comunicazione, e le problematiche connesse allo sfruttamento dell'output basato sui registri e la loro integrazione.

La nuova infrastruttura informativa del Framework RAF assume a riferimento basi di dati caratterizzate dall'assenza di un "disegno d'indagine", dal carattere di esaustività dei dati individuali rispetto a intere popolazioni e dalla integrabilità dei dati elementari, sulla base delle regole definite dal Sistema Integrato dei Registri (SIR). Queste proprietà delimitano il perimetro del progetto, orientandolo strettamente ai registri statistici ed escludendo basi dati con caratteristiche non coerenti con i requisiti espressi.

La particolare natura delle basi dati trattate dal progetto e gli obiettivi attesi in termini di impatto informativo rendono necessario prefigurare alcuni orientamenti di fondo:

- la diversa tipologia di registro, base, esteso e tematico, unitamente alle sue caratteristiche (grado di multidimensionalità, compresenza di diverse unità statistiche ecc.) prefigura un progressivo aumento della complessità informativa disponibile. In generale si suppone che la nuova infrastruttura RAF potrà presentare un'operatività più immediata sui registri base, caratterizzati da molte unità e poche variabili; un'operatività di complessità crescente sui registri estesi e sui registri tematici; un'operatività di elevata complessità sul SIR, considerando che appaiono particolarmente vaste le possibilità di incrocio sulla base delle diverse classificazioni statistiche adottate.
- L'interazione tra requisiti tematici, tecnologici e potenza di calcolo è caratterizzata da un forte dinamismo: la dimensione tecnologica appare cruciale per la riuscita del progetto, in relazione sia alla dimensione quantitativa delle informazioni presenti nei registri, sia, soprattutto, alla gestione delle componenti di maggiore complessità degli output ipotizzati.
- L'elevato potenziale informativo del sistema richiede la definizione di moduli informativi differenziati sulla base dell'analisi della segmentazione della domanda d'informazione statistica e di una efficace classificazione delle diverse tipologia di utenza, in termini sia di contenuti informativi sia di modalità di fruizione. La rilevanza dei risultati del progetto trova nelle modalità di interazione tra offerta e domanda un fattore critico.
- Il Framework dovrà essere caratterizzato da una piena scalabilità, in modo da soddisfare le specifiche esigenze di differenti classi e livelli di utenza, sia interna sia esterna. Dovrà essere previsto, ad esempio, l'accesso ai microdati a livello di sistema integrato dei registri per alcune classi particolari di utenza, l'accesso ai dati elementari a livello di singolo registro di base secondo quanto previsto nel PSN per, ad esempio, gli uffici del Sistan, fino all'accesso ai soli dati aggregati per le classi di utenza più generiche. La scalabilità nello sfruttamento del sistema dovrà tenere conto della combinazione di una molteplicità di fattori, quali, ad esempio, specifiche policy di abilitazione degli utenti; effettive capacità di richiesta e analisi dei dati; natura dei tematismi e diverse possibilità di integrazione.
- La componente micro dell'informazione disponibile deve essere adeguatamente valorizzata, se classificabile come componente "core" del singolo registro da trattare, o del SIR. Questa caratterizzazione rende particolarmente efficaci policy di diffusione basate, ad esempio, su microaggregazioni dei dati, pianificate o richieste dall'utente.
- Nella prima fase progettuale è prevista la realizzazione di un prototipo, che assumerà a riferimento, da un punto di vista tematico, il registro Frame SBS territoriale e il registro del Lavoro in considerazione del grado di maturità di

costruzione degli stessi ma dovrà essere generalizzabile agli altri registri e alla loro integrazione.

Come ulteriore elemento strategico va sottolineato che il Framework RAF si inserisce appieno nelle politiche di diffusione dell'Istituto, affiancando ed integrando le altre piattaforme. In particolare, I.Stat continuerà a diffondere i dati derivanti da indagini tradizionali che per natura peculiare non possono essere sostituite dai registri e che quindi continuano ad avere una vita autonoma rispetto a questi ultimi. Analogamente, la piattaforma Linked Open Data – anche in prospettiva del suo nuovo impianto architeturale – mantiene il suo ruolo centrale nella diffusione di dati aperti, destinato in particolare al rilascio di dati a livello subcomunale non diffusi attraverso I.Stat .

4 Tutela della privacy e impatto sulla statistica ufficiale

Dal 25 maggio del 2018, il nuovo regolamento generale per la protezione dei dati, regolamento (UE) n. 2016/679, meglio noto come GDPR, è diventato operativo. Tale regolamento sostituisce la direttiva europea sulla protezione dei dati 95/46/EC con l'obiettivo di armonizzare le norme per la protezione dei dati personali nell'unione Europea. Il GDPR e la recente approvazione del d.lgs. n. 101/2018, che aggiorna la preesistente normativa nazionale in materia contenuta nel d.lgs. n. 196/03 (codice in materia di protezione dati personali), hanno portato un notevole cambiamento nella gestione della privacy la cui applicazione produce un notevole impatto sui processi statistici sia per la produzione che per la ricerca. A seguito dell'adeguamento e della trasformazione del codice di deontologia (allegato A3 del d.lgs. n. 196/03) - denominato Regole deontologiche per trattamenti a fini statistici o di ricerca scientifica effettuati nell'ambito del Sistema Statistico nazionale - effettuato dal Garante per la protezione dei dati personali (prov. 19 dicembre 2018), l'Istat ha promosso la creazione di un tavolo dei Responsabili Protezione Dati in ambito Sistan che si occuperà di comprendere come declinare in ambito statistico le prescrizioni presenti nel nuovo regolamento e di produrre una nuova versione aggiornata delle Regole deontologiche.

L'obiettivo sarà di preservare le specificità dei trattamenti a fini statistici effettuati dai soggetti appartenenti al Sistan, in modo da consentire di rispondere efficacemente alla crescente domanda di analisi che l'attuale società dell'informazione pone in modo sempre più stringente alla statistica.

L'introduzione del principio di *accountability* per il titolare (art. 24 GDPR), unitamente alla presenza di un severo schema sanzionatorio (art. 83 GDPR), costituiscono i principali punti di innovazione presenti nel nuovo regolamento.

La privacy del dato trattato deve essere considerata come fattore abilitante per l'organizzazione della statistica ufficiale fin dalle prime fasi di progettazione secondo il

paradigma della privacy by design (art. 25 GDPR). Questo comporta la necessità di un cambio culturale che impone di prestare attenzione alla protezione dei dati nell'intero di ciclo di vita del dato, dalle scelte metodologiche alla conservazione o distruzione del dato trattato, documentando le scelte effettuate e valutando i possibili rischi nonché l'impatto potenziale sui diritti e sulle libertà degli interessati. A tal proposito devono essere sempre tenuti in considerazione i seguenti principi come descritti nell'art. 5 GDPR:

- a) liceità, correttezza e trasparenza, per trattare i dati bisogna sempre individuare una base giuridica che ne legittimi il corretto utilizzo e le modalità di comunicazione in modo trasparente agli interessati;
- b) limitazione delle finalità, i dati devono essere raccolti per finalità determinate, esplicite e legittime, e successivamente trattati in modo che non sia incompatibile con tali finalità (la finalità statistica, unitamente a quanto indicato nell'art. 89 GDPR, non è incompatibile come finalità per trattamento ulteriore dei dati)
- c) minimizzazione, i dati trattati devono essere sempre adeguati, pertinenti e limitati a quanto necessario per il perseguimento delle finalità per cui sono raccolti e trattati;
- d) esattezza, devono essere adottate tutte le misure ragionevoli per cancellare o rettificare tempestivamente i dati inesatti rispetto alle finalità per le quali sono trattati;
- e) limitazione della conservazione, i dati devono essere 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 per fini statistici - o per quanto indicato nell'art. 89 GDPR - fatta salva l'attuazione di misure tecniche e organizzative adeguate richieste dal presente regolamento a tutela dei diritti e delle libertà dell'interessato);
- f) integrità e riservatezza, i dati devono essere trattati in maniera da garantire un'adeguata sicurezza dei dati personali, compresa la protezione, mediante misure tecniche e organizzative adeguate, da trattamenti non autorizzati o illeciti e dalla perdita, dalla distruzione o dal danno accidentali.

Oltre ai principi indicati il regolamento prescrive alcune azioni come obbligatorie per la conformativa normativa, le principali sono:

- a) la creazione e il mantenimento di un registro dei trattamenti che elenchi tutte le attività che entrano in contatto con dati personali, descrivendone le caratteristiche in modo da consentire un monitoraggio che permetta di individuare in ogni momento il personale incaricato del trattamento e le misure di sicurezza adottate;

- b) la nomina di un Responsabile Protezione Dati (DPO) e di una struttura a supporto del titolare del trattamento dati che mantenga i rapporti con l'autorità garante della privacy, operando in autonomia e con la disposizione di risorse adeguate al compito da svolgere.
- c) l'effettuazione di un documento di valutazione impatto privacy laddove si evidenzia un potenziale "rischio elevato per i diritti e le libertà delle persone fisiche" ai sensi dell'art. 35 GDPR.

Le attività di adeguamento che il nuovo quadro normativo impone congiuntamente al cambiamento culturale già indicate sono da ricercarsi nella individuazione di soluzioni:

- organizzative che permettano la chiara gestione delle responsabilità di tutti i soggetti coinvolti nel trattamento dei dati personali all'interno o all'esterno dell'Istituto e dei rapporti con gli interessati, il costante aggiornamento del registro dei trattamenti, la produzione di valutazioni di impatto privacy per le varie attività di trattamento;
- tecnologiche che permettano di progettare e sviluppare gli applicativi statistici secondo i principi di privacy by design e by default, facendo uso di tecniche di pseudonimizzazione dati e procedure efficienti per la gestione di data breach;
- formative per accrescere presso il personale dell'Istituto la conoscenza del nuovo regolamento e gli adempimenti correlati.