

VERSIONE PROVVISORIA



**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**

Commissione per la garanzia della qualità dell'informazione statistica

Roma, 22 settembre 2017

Indice

1	Inquadramento normativo	5
2	La gestione dei microdati di fonte amministrativa	7
	<i>2.1 I metodi di integrazione dei dati amministrativi: il Sistema Integrato di Microdati SIM</i>	<i>8</i>
	<i>2.2 La documentazione della qualità dei dati amministrativi</i>	<i>10</i>
	<i>2.3 I dati amministrativi a supporto dei Censimenti Permanenti</i>	<i>11</i>
3	Miglioramenti nelle politiche di accesso ai microdati	13
	<i>3.1 Il Laboratorio ADELE e l'accesso da remoto</i>	<i>13</i>
	<i>3.2 Laboratorio ADELE: ulteriori innovazioni</i>	<i>17</i>
	<i>3.3 L'informazione sui microdati</i>	<i>19</i>

1 Inquadramento normativo

Un corretto inquadramento normativo delle attività di raccolta, utilizzo e conservazione dei microdati all'interno del Sistan non può non partire dalla constatazione che l'attuale assetto è interessato da forti cambiamenti.

Il Disegno di legge n. 4620 (già approvato dal Senato e ora all'esame della 14ª Commissione permanente della Camera dei deputati) delega, infatti, il Governo ad adottare, entro sei mesi dalla data di entrata in vigore della legge, uno o più decreti legislativi al fine di adeguare il quadro normativo nazionale alle disposizioni del Regolamento (UE) n. 2016/679. Gli obiettivi sono molteplici: abrogare le disposizioni del d. lgs. n. 196/2003 non compatibili con le disposizioni del regolamento europeo; modificare il decreto laddove necessario per dare attuazione alle disposizioni non direttamente applicabili contenute nel regolamento; coordinare le disposizioni vigenti in materia di protezione dei dati personali con le disposizioni del regolamento; prevedere, ove opportuno, il ricorso a specifici provvedimenti attuativi e integrativi del Garante per la protezione dei dati personali e, infine, adeguare il sistema sanzionatorio vigente alle disposizioni del regolamento.

A distanza di venti anni dall'entrata in vigore della prima legge sulla protezione dei dati personali (L. n. 675/1996) e a quindici dall'emanazione del Codice deontologico per il Sistan (Provvedimento del Garante n. 13 del 31 luglio 2002), siamo dunque nuovamente chiamati a contribuire attivamente affinché le garanzie per la riservatezza degli interessati si coniughino al meglio con la capacità del Sistema di fornire risposte adeguate alle istanze di cambiamento che in maniera sempre più pressante investono il governo della statistica ufficiale ma, al contrario, siano occasione per aggiungere valore a ciò che facciamo.

Si è più volte osservato, anche in questa sede istituzionale, che la legislazione vigente contiene una corposa serie di deroghe per il Sistan, che hanno consentito di conciliare efficienza e tutela dei diritti nell'ambito di un giusto bilanciamento tra i diversi interessi in gioco.

Tuttavia, a fronte del crescente *favor* dell'ordinamento nazionale ed europeo per l'utilizzo di fonti di dati alternative alle rilevazioni dirette, finalizzato principalmente accrescere la tempestività e la qualità della produzione statistica, a ridurre l'onere sui rispondenti e a contribuire all'uso efficiente delle risorse disponibili per la produzione dell'informazione statistica e in una fase storica in cui l'Istat è impegnato nella realizzazione del sistema integrato di registri statistici attraverso l'utilizzo di fonti da indagine, fonti amministrative e nuove fonti, non possiamo non segnalare le criticità derivanti da alcuni limiti all'uso di dati personali nell'integrazione di fonti, limiti evidenziati in particolare negli ultimi due pareri resi dal Garante della privacy (*PSN 2014-2016 aggiornamento 2015-2016* e *PSN 2017-2019*). Problemi non trascurabili sono generati, in particolare, dalla necessità di mantenere l'utilizzo dei dati reperibili liberamente *on line* e più in generale delle grandi aggregazioni di dati (*Big data*), all'interno di un quadro di regole che, evidentemente, era stato costruito senza tener conto dell'impatto che tale enorme disponibilità di dati avrebbe avuto sulla conoscenza e l'informazione statistica, in particolare.

Va detto che l'Istat si è doverosamente attenuto alle prescrizioni del Garante sospendendo talune operazioni di trattamento in attesa della programmata *Verifica preliminare sul Sistema di integrazione logico-fisica di microdati amministrativi e statistici (SIM)*, fornendo una documentata descrizione di metodi e tecniche utilizzati in via sperimentale, in particolare nell'ambito dei progetti previsti nella scheda IST-02589 "*Uso a fini statistici dei Big Data*", e adottando le ulteriori specifiche cautele a tutela degli interessati sollecitate dal Garante.

Allo stesso tempo, però, è emersa la necessità di attivare idonee iniziative affinché nell'elaborazione delle nuove regole nazionali siano valorizzati gli spazi di flessibilità rinvenibili nel Regolamento Europeo. Abbandonata la logica del mero adempimento formale, tale Regolamento prospetta, infatti, un chiaro approccio *risk based*: responsabilizzazione del Titolare nella rilevazione/gestione dei rischi per i diritti degli interessati, in funzione di una reale protezione dei dati personali. Ciò si traduce nell'obbligo di valutare preventivamente l'effettiva attitudine al danno di ogni trattamento e adottare adeguate "contromisure". E allora, in quest'ottica sarebbe opportuno puntare al rafforzamento della capacità dell'ordinamento statistico di darsi ulteriori regole responsabilmente cosicché il rilascio di un'informazione di qualità, tempestiva e sempre più dettagliata, non precluda il pieno rispetto della privacy delle persone fisiche e più in generale del segreto statistico. Certamente l'Istat, per la sua storia e per le esperienze maturate, anche in campo internazionale, è nelle condizioni di

fornire precise indicazioni al riguardo. In questa prospettiva i codici deontologici potranno continuare a giocare un ruolo fondamentale, per giungere a un nuovo bilanciamento degli interessi coinvolti.

Allo stesso tempo dobbiamo preservare e consolidare le disposizioni vigenti che, come si è detto, riconoscendo la specificità dei trattamenti effettuati per il perseguimento degli scopi della statistica ufficiale, semplificano importanti adempimenti (una fra tutte l'informativa agli interessati quando i dati sono raccolti presso terzi, obbligo che si considera assolto con la descrizione nel PSN delle caratteristiche del trattamento).

Mentre, dunque, è in cantiere la riscrittura delle norme sull'utilizzo dei microdati all'interno del processo di produzione della statistica ufficiale, il rilascio all'esterno (comunicazione e diffusione) è sufficientemente coperto a livello normativo, anche grazie alla recente nuova disciplina in materia di Accesso per fini scientifici ai dati elementari raccolti per finalità statistiche, disciplina che nasce con il decreto legislativo 25 maggio 2016 n. 97 (che ha introdotto l'art. 5-ter nel d.lgs.n. 33/2013) e si completa con le Linee guida del Comstat, attualmente al vaglio del Garante della privacy.

Una volta a regime le nuove norme permetteranno di consolidare il partenariato con soggetti esterni al Sistan, con la comunità scientifica e le università in particolare. Si tratta di un sensibile avanzamento sul piano normativo a cui deve seguire il rafforzamento della capacità di tradurre l'accresciuta fruibilità di microdati in concreti benefici per la conoscenza.

2 La gestione dei microdati di fonte amministrativa

Nell'ambito del processo di modernizzazione della statistica ufficiale, incentrato prevalentemente sulla implementazione di un sistema integrato di registri statistici, l'utilizzo dei microdati di fonte amministrativa a fini statistici assume una particolare rilevanza e rappresenta una delle linee strategiche dell'Istituto.

L'Istat, infatti sta perseguendo il progetto che prevede il passaggio da un modello «tradizionale» di rilevazione, basato sull'acquisizione diretta dei dati dalle fonti (cittadini e imprese), ad un modello basato su un sistema integrato di registri statistici, derivante dall'utilizzo di fonti da indagine, fonti amministrative e nuove fonti (ad esempio i cosiddetti Big data).

In questo nuovo scenario, la politica di gestione dei microdati di fonte amministrativa da parte dell'Istituto deve tener conto da una parte delle esigenze degli utilizzatori, in

primis della comunità scientifica e dei produttori di statistiche, e dall'altra dei vincoli e delle regole dettati dalla normativa in materia di protezione dei dati personali.

2.1 I metodi di integrazione dei dati amministrativi: il Sistema Integrato di Microdati SIM

Il Sistema Integrato di Microdati (SIM) su individui, famiglie e unità economiche è un REPOSITORY di microdati (dati elementari) integrati costruito a partire da archivi di fonte amministrativa. Il SIM è pensato come un ambiente sicuro per la conservazione dei dati amministrativi. La sua realizzazione consente al tempo stesso di effettuare controlli di qualità sull'identificazione delle unità contenute negli archivi per garantirne la validità statistica e di procedere a integrazioni logiche di dati mediante l'utilizzo di tecniche di record linkage che attribuiscono un codice numerico "non parlante" alle unità contenute negli archivi.

La funzione di acquisizione, integrazione e rilascio dei dati amministrativi per finalità statistiche è stata completamente centralizzata all'interno della Direzione che si occupa della raccolta dati. La programmazione dell'acquisizione degli archivi amministrativi viene effettuata annualmente secondo le seguenti fasi:

- a) raccolta delle esigenze informative presso le Direzioni tecniche e i servizi di produzione statistica;
- b) analisi congiunta delle richieste con le Direzioni tecniche e i servizi di produzione;
- c) verifica della conformità alla normativa vigente in materia di acquisizione e utilizzo dei dati amministrativi per uso statistico;
- d) nel caso di dati personali verifica del rispetto delle norme sulla protezione dei dati personali. La normativa vigente, infatti, consente l'acquisizione dei dati amministrativi solo per le finalità statistiche indicate nel Programma statistico nazionale in corso di validità o per le finalità statistiche connesse ad adempimenti derivanti da Regolamenti Europei, ovvero per altre finalità statistiche legate a progetti di ricerca di particolare rilevanza. La pubblicazione in gazzetta ufficiale del PSN assolve alla funzione di informativa per i soggetti interessati, in conformità con quanto previsto dall'art. 6 del Codice di deontologia e di buona condotta per i trattamenti di dati personali a scopi statistici e di ricerca scientifica effettuati nell'ambito del Sistema statistico nazionale, e per i dati sensibili e/o giudiziari assolve anche la funzione di autorizzazione al trattamento.
- e) definizione di una prima bozza di programmazione annuale delle acquisizioni;
- f) condivisione della prima bozza di programmazione con i vari Enti titolari delle fonti amministrative, attraverso riunioni dedicate in cui vengono analizzate

dettagliatamente le richieste con riguardo a fattibilità, tempistica delle forniture, periodo di riferimento dei dati, modalità tecniche etc.

In alcuni casi sono stipulati appositi protocolli d'intesa per la collaborazione in ambito statistico.

L'acquisizione da parte dell'Istat dei dati amministrativi che alimentano il SIM avviene attraverso un portale unico di raccolta informatica dei dati di titolarità di enti pubblici e privati denominato "Sistema di acquisizione degli archivi amministrativi - ARCAM" (<https://arcam.istat.it>). L'utilizzo di tale Portale rappresenta l'esito di un processo di reingegnerizzazione del sistema di acquisizione dei dati. Un processo fortemente innovativo attuato nel corso del biennio 2015/2016 e finalizzato a garantire una maggiore sicurezza nella trasmissione delle informazioni, l'integrità dei flussi di dati verso l'Istituto e l'utilizzo di tecnologie standard conformi ai requisiti prescritti dalla legislazione in materia di trattamento dei dati personali (per un approfondimento sul funzionamento del portale ARCAM si rimanda all'Allegato 1).

L'archiviazione dei dati nell'ambito del Sistema SIM si avvale di un DataBase di tipo relazionale e prevede la eliminazione dei dati identificativi diretti delle unità statistiche (persone fisiche e unità economiche) dall'archivio di base e la loro conservazione separata. A ciascuna unità statistica viene attribuito, come accennato, un codice numerico - denominato Codice SIM - "non parlante" (pseudonimizzazione) che consente il collegamento tra le stesse unità statistiche presenti negli archivi contenuti nel Sistema.

I dati identificativi delle unità statistiche e il relativo Codice SIM sono conservati in strutture di dati memorizzate in schemi distinti del database, separati quindi dalle altre informazioni presenti nel SIM. A questi schemi sono associate specifiche credenziali di accesso attribuite a un numero ristretto di soggetti incaricati della procedura di implementazione e gestione del Sistema (per un approfondimento sui metodi di integrazione dei dati amministrativi si rimanda all'Allegato 2).

I dati contenuti nel Sistema SIM, sono resi accessibili sotto forma di viste parziali prive dei campi di non interesse e in forma pseudonimizzata al personale dell'Istat, con utenze nominali, espressamente incaricato del loro trattamento ai sensi del decreto legislativo n. 196 del 2003 per la realizzazione dei lavori statistici dell'Istat.

Il responsabile del trattamento dei dati personali è il Direttore centrale della raccolta dati, a cui l'interessato si può rivolgere per l'esercizio dei propri diritti.

I dati presenti nel SIM si riferiscono a tre tipi di unità d'interesse statistico (gli individui, le unità economiche e i luoghi) e a due tipi di unità derivate dalla identificazione delle predette unità statistiche, (i luoghi associati agli individui e alle unità economiche e le relazioni tra individui e unità economiche).

I dati sensibili presenti in alcuni degli archivi presenti nel SIM sono conservati in schemi di database separati dai dati identificativi; in particolare, le variabili sensibili sono contenute in tabelle accessibili, mediante user-id e password, esclusivamente dal responsabile del trattamento dei dati personali del Sistema SIM e dal responsabile del servizio incaricato della gestione del medesimo Sistema.

I dati sensibili sono trattati mediante l'utilizzazione di soluzioni che, considerato il numero e la natura dei dati trattati, li rendono temporaneamente inintelligibili anche a chi è autorizzato ad accedervi e permettono di identificare gli interessati solo in caso di necessità, conformemente a quanto prescritto dall'art. 22, commi 6 e 7 del decreto legislativo n. 196 del 2003.

L'accesso alle variabili sensibili da parte del personale dell'Istat, a seguito di procedura interna di accreditamento, è consentito esclusivamente ai soggetti incaricati del trattamento di dati personali, per la realizzazione di progetti inseriti nel Programma statistico nazionale per i quali sia stato dichiarato l'utilizzo di dati sensibili nella relativa scheda identificativa, ovvero di trattamenti per i quali sussistano i presupposti prescritti dall'art. 20 del d.lgs. n. 196 del 2003 (per un approfondimento sulle procedure rilascio di dati amministrativi agli utilizzatori interni si rimanda all'Allegato 3)

2.2 La documentazione della qualità dei dati amministrativi

Un'importante innovazione in tema di supporto ai processi di produzione statistica a cui sta lavorando l'Istituto è la predisposizione della QRCA, acronimo di "Quality Report Card dei dati Amministrativi".

Il sistema permetterà agli utenti interni di avere informazioni e documentazione per il corretto utilizzo di ogni archivio amministrativo acquisito da enti esterni (tracciati record, contenuti, anni disponibili, descrizione del processo di integrazione, indicatori di qualità, ecc.).

Il framework della qualità utilizzato nella QRCA è stato sviluppato nell'ambito del progetto internazionale BlueEts, finanziato dalla Commissione Europea.

Si tratta di un progetto, in via di implementazione, basato sull'interoperabilità dei sistemi che trattano dati amministrativi, in particolare il portale ARCAM, il sistema SIM e il PSN (Programma Statistico Nazionale).

2.3 I dati amministrativi a supporto dei Censimenti Permanenti

Il passaggio dai tradizionali censimenti decennali esaustivi ai censimenti permanenti prevede l'utilizzo congiunto di dati provenienti da archivi amministrativi e di quelli rilevati da altre indagini campionarie dell'Istituto, oltre all'utilizzo di nuove tipologie di fonti come i Big Data. Il nuovo sistema dei censimenti permanenti contribuirà, in coerenza con le finalità censuarie, alla realizzazione di un nuovo sistema di Registri di Base integrati (della popolazione, delle unità economiche e delle istituzioni, dei luoghi e delle attività).

In particolare, nel caso dei dati sulla popolazione il passaggio dal censimento decennale al censimento permanente rappresenta un elemento di profonda innovazione, con una ingente riduzione dei costi dell'operazione e un aumento della frequenza di produzione e diffusione dei dati. Ciò sarà importante anche in riferimento alla funzione di Vigilanza anagrafica che l'Istat è tenuto a svolgere insieme al Ministero dell'Interno. In tale ambito, infatti, la frequenza annuale della produzione dei dati censuari avrà notevoli effetti sulla possibilità di monitorare con continuità la qualità delle Anagrafi (in prospettiva ANPR) e, conseguentemente, garantirà maggiore stabilità all'andamento del calcolo della popolazione (con riduzione pressoché totale della necessità di ricorrere a ricostruzioni delle serie, prima necessariamente legate alla cedenza decennale della produzione della popolazione legale). Questo passaggio è reso possibile proprio grazie alla ricchezza informativa dei dati amministrativi già acquisiti e integrati.

Tale patrimonio è in crescita proprio per supportare al meglio la nuova strategia censuaria.

Sono infatti previste due attività specifiche:

- la memorizzazione e l'integrazione in SIM degli altri archivi su individui, unità economiche e luoghi già correntemente acquisiti (anche tramite indagini);
- l'acquisizione e l'integrazione in SIM di nuovi archivi ad alto potenziale informativo come ad esempio i dati sui consumi energetici delle famiglie e delle imprese e le comunicazioni obbligatorie riguardanti i rapporti di lavoro.

2.4 L'investimento sulla qualità dei microdati raccolti tramite indagini

La centralizzazione della raccolta dati prevista dal disegno di modernizzazione dell'Istituto ha favorito l'attivazione di funzioni trasversali dedicate alla standardizzazione e armonizzazione dei processi di acquisizione dei microdati di indagine, contribuendo al miglioramento dei dati raccolti presso le unità di rilevazione. L'attenzione alla qualità dei dati viene perseguita con l'applicazione di rigorose e innovative metodologie di raccolta dati e con costante attenzione alla necessità di coniugare la riduzione dei costi e del burden sul rispondente con il miglioramento della qualità dei dati raccolti.

I filoni di attività nei quali si è investito soprattutto negli ultimi anni sono diversi. Innanzitutto, le opportunità offerte dalle nuove tecnologie hanno ispirato la transizione a tecniche computer assisted per la realizzazione delle indagini su imprese e famiglie consentendo di migliorare la tempestività del rilascio delle informazioni statistiche e la riduzione dell'errore non campionario. In particolare, il crescente ricorso alle tecniche miste per la rilevazione dei dati ha consentito di migliorare la copertura della popolazione target e la riduzione dei costi, accrescendo al contempo l'efficienza dell'intero processo di raccolta dati.

Parallelamente, si è lavorato per la razionalizzazione e armonizzazione degli strumenti di rilevazione (questionari). Lo sfruttamento di dati da fonte amministrativa, la riduzione delle ridondanze e la costruzione di questionari sempre più user friendly hanno consentito di ridurre il burden sul rispondente.

L'armonizzazione dei quesiti e l'adozione di classificazioni standardizzate per l'acquisizione di informazioni comuni a più rilevazioni sono modalità operative ormai generalizzate che favoriscono l'integrazione e la comparabilità dei dati raccolti.

La realizzazione di cognitive test e il monitoraggio di tutte le fasi del lavoro sul campo rappresentano un modus operandi ormai consolidato che ha consentito di migliorare, affinandola, la formulazione dei quesiti e la sequenza degli stessi, garantendo una più puntuale misurazione dei fenomeni oggetto di interesse.

Anche le innovazioni introdotte nelle strategie di formazione e monitoraggio delle reti di rilevazione, pensate sempre in un'ottica di armonizzazione e di diffusione delle best practice, hanno contribuito alla produzione di un dato statistico di qualità elevata.

3 Miglioramenti nelle politiche di accesso ai microdati

La disponibilità di basi di microdati risultanti da indagini e fonti amministrative è un aspetto cruciale per l'avanzamento della ricerca scientifica. In tutti i paesi industrializzati gli istituti nazionali di statistica sono chiamati a operare per fornire ai ricercatori, ai policy maker e alla società tutta, dati di qualità, quanto più possibile analitici e pertinenti, sia per analisi a carattere tematico (economico, sociale, ambientale, ecc.), sia per analisi territoriali.

Per rispondere efficacemente a questo obiettivo, è imprescindibile un dialogo intenso tra Istat e comunità scientifica, che consenta di verificare continuamente l'adeguatezza dei dati e degli strumenti interpretativi messi a disposizione del Paese, ed eventualmente aggiornarli. La sfida è permettere a ciascuno, istituzione o cittadino, di operare scelte informate e di mettere gli studiosi e i ricercatori, in particolare, nelle migliori condizioni per offrire il loro contributo di idee ed elaborare nuove prospettive di analisi.

3.1 Il Laboratorio ADELE e l'accesso da remoto

L'Istat, cogliendo lo stimolo che viene dall'Unione europea, si è impegnato molto, nei limiti accordati dalle norme vigenti¹, per ampliare quanto più possibile l'utilizzo dei microdati.

Dal 2013 ha iniziato la produzione dei file mlcro.STAT, file a uso pubblico direttamente scaricabili dal sito web dell'Istituto e dal 2016 ha messo a disposizione un tipo particolare di file a uso pubblico, i "mlcro.STAT storici", grazie ai quali anche un pubblico non specialista può liberamente analizzare le indagini maggiormente richieste, anche all'indietro nel tempo per circa 8-10 anni: Rilevazione sulle forze di lavoro (2008-2016), Aspetti della vita quotidiana (2005-2012) e Indagine sui consumi delle famiglie (2004-2013).

L'offerta di microdati ha anche visto un costante incremento del numero di file per la ricerca. A oggi sono stati creati file MFR per 18 indagini, di cui 3 nell'ultimo anno: Condizione e integrazione sociale dei cittadini stranieri, Multiscopo sulle famiglie: uso del tempo, Consumi energetici delle famiglie.

¹ Art. 10, comma 2, del d.lgs. 322/1989 e art. 7 del Codice deontologico per i soggetti Sistan.

Dal 2016, sono stati messi a disposizione del Laboratorio ADELE 7 nuovi file risultanti dall'integrazione di dati provenienti da più rilevazioni (file integrati), mentre cinque sono stati aggiornamenti di edizioni precedenti, a conferma di come l'investimento fatto in questa direzione stia progressivamente entrando a regime.

Sempre nel corso del 2016, sono state riviste e ampliate le regole per il controllo finale dell'output, permettendo ai ricercatori di condurre nel laboratorio ADELE analisi che in precedenza non erano consentite (Equazioni strutturali, Analisi fattoriale, Analisi delle corrispondenze, Analisi in componenti principali).

Malgrado i risultati raggiunti nell'offerta per l'utilizzo dei microdati, persistevano però alcune criticità derivanti per lo più dall'impianto normativo, criticità che anche Eurostat aveva rilevato nel corso della *peer review* condotta presso il nostro Istituto nel 2015 e conclusa con la raccomandazione, tra le altre, di offrire anche in Italia un accesso ai microdati da remoto, in coerenza col Regolamento Ue n. 557/2013.

Nel corso del 2016, l'Istat ha quindi sollecitato un'azione legislativa, sfociata nel Decreto Legislativo n. 97/2016 (per il testo del decreto si rimanda all'Allegato 4)

Il Decreto Legislativo n. 97/2016 (d'ora in avanti D.lgs.) segna un deciso avanzamento. La nuova normativa, nel solco del Regolamento (Ue), definisce il quadro legale in cui incardinare l'accesso ai microdati del Sistan a uso scientifico, con riferimento ai file per la ricerca (MFR), ai laboratori per l'accesso ai dati elementari e ai protocolli di ricerca. Più in generale, pone le basi giuridiche necessarie per dare avvio alle conseguenti attività organizzative e tecnologiche.

I principali contenuti del nuovo D.lgs. sono già stati oggetto di trattazione nella audizione dello scorso febbraio. Qui, si ricordano due dei principali tratti innovativi delle nuove disposizioni.

Il primo riguarda l'ampliamento del novero dei possibili fruitori dei microdati, che viene esteso al di là dei ricercatori appartenenti a università o istituti di ricerca. In base al D.lgs. infatti possono essere riconosciuti come enti di ricerca, aventi diritto all'accesso ai dati elementari a fini scientifici, anche singole strutture – quali i dipartimenti – interne ad amministrazioni o enti pubblici e privati, sempre che svolgano attività di ricerca. Vengono così superate alcune limitazioni che finora avevano portato ad escludere, ad esempio, gli uffici studi di banche, di organizzazioni di categoria eccetera.

La prospettiva più rilevante che si apre con il D.lgs. è comunque quella di "avvicinare" il Laboratorio ADELE ai ricercatori, per facilitare quanti – per le loro analisi –

necessitano di elaborare direttamente le variabili originarie senza limitazioni di sorta, se non per quanto attiene alle variabili identificative, sensibili e giudiziarie².

L'obiettivo è quello di costituire anche in Italia un accesso da remoto ai microdati, organizzando, presso alcuni centri particolarmente attivi nell'analisi quantitativa, dei *safe centre* dai quali sia possibile accedere in modo sicuro al patrimonio informativo dell'istituto. Anche il mondo della ricerca italiano, così come già accade per i principali istituti di ricerca europei, avrà così l'opportunità di sfruttare le potenzialità di analisi proprie del Laboratorio, senza l'obbligo di recarsi fisicamente presso uffici Istat (o di altri enti Sistan), siano essi dislocati presso la sede centrale di Roma o presso gli uffici territoriali.

Il D.lgs. prevede che affinché queste innovazioni possano essere poste in essere, il COMSTAT, con il supporto tecnico dell'Istat, elabori Linee guida che definiscano:

- a) i criteri per il **riconoscimento degli enti** che possono accedere ai microdati, sulla base degli scopi istituzionali derivanti dall'atto costitutivo/statuto, dell'attività di ricerca svolta e della reputazione scientifica, dell'organizzazione interna in relazione all'attività di ricerca accertando, in particolare, la sussistenza di sufficienti condizioni di autonomia, delle misure tecniche e delle infrastrutture adottate per garantire la sicurezza dei dati;
- b) i criteri di **ammissibilità dei progetti di ricerca**, valutando lo scopo scientifico della ricerca, l'effettiva necessità di disporre dei dati richiesti, i risultati attesi e i benefici che possono derivare dalla loro pubblicazione, l'esperienza e la professionalità dei ricercatori coinvolti e i metodi impiegati per la loro analisi e diffusione, avuto riguardo anche alla tutela del segreto statistico e alla protezione dei dati personali;
- c) le modalità di **organizzazione dei laboratori** e i criteri per l'accertamento **dell'idoneità dei soggetti cui affidare la gestione** dei laboratori virtuali, valutando gli scopi istituzionali perseguiti, la reputazione di cui il soggetto gode, l'adeguatezza della struttura organizzativa in relazione all'attività da svolgere, la competenza ed esperienza del personale addetto al servizio, le misure tecniche e procedurali previste a garanzia di un'adeguata protezione, fisica e logica, dei dati;
- d) le **sanzioni applicabili in caso di violazione** degli impegni assunti dall'ente e dai singoli ricercatori, quali: inibire al ricercatore l'accesso ai dati, richiedere all'istituzione di appartenenza l'adozione di provvedimenti disciplinari nei suoi

² Le uniche limitazioni alle informazioni imposte nel Laboratorio ADELE riguardano gli identificativi diretti e le variabili sensibili e giudiziarie che vengono tutti rimossi.

confronti, effettuare segnalazioni agli organi di polizia, attivare iniziative legali a tutela dell'interesse del titolare dei dati.

Tra ottobre 2016 e marzo 2017 Istat e Comstat hanno lavorato intensamente a questo obiettivo e ad aprile 2017 le Linee guida sono state inviate al Garante per la protezione dei dati personali, per il previsto parere.

Per quanto riguarda l'accesso da remoto ai microdati le Linee guida³ prevedono che ciascun ente Sistan che intende mettere a disposizione questa modalità di fruizione possa svolgere una sperimentazione per valutare l'infrastruttura tecnologica e le misure per la sicurezza dei dati più adeguate, nonché l'onere organizzativo e finanziario che l'accesso da remoto comporta.

La sperimentazione può essere condotta in collaborazione con un soggetto individuato sulla base di specifici criteri tra i quali le sue caratteristiche scientifiche, tecnologiche e logistiche, l'utenza potenziale del laboratorio, le misure messe in atto per la sicurezza dei dati.

Sulla base dei criteri appena esposti, Istat ha recentemente individuato come possibili partner il Dipartimento di Metodi e Modelli per l'Economia, il Territorio e la Finanza dell'Università La Sapienza e il Dipartimento di economia e statistica della Banca d'Italia.

Dopo i necessari contatti, qualora sussistessero tutte le condizioni previste dalle norme, l'Istat e i due enti potranno sottoscrivere appositi accordi e avviare la fase pilota.

La sperimentazione riguarda tanto gli aspetti informatici relativi alla sicurezza dei dati (configurazione della rete e dei pc, sistemi di autenticazione, eccetera), quanto quelli organizzativi e procedurali: il presidio del *safe centre*, la gestione delle richieste, la valutazione e il rilascio degli output.

I risultati dell'esperimento consentiranno di valutare sia a chi gestisce il servizio sia ad Istat quali siano le soluzioni più adatte da realizzare, una volta a regime.

³Si segnala che le Linee guida, nell'ambito della sezione dedicata ai file MFR, prevedono la possibilità per la Commissione per la garanzia dell'informazione statistica di richiedere agli enti del Sistan produttori di questo tipo di file la documentazione relativa alle misure di protezione adottate, misure che – sulla base dei protocolli consolidati in materia di riservatezza – non vengono rese pubbliche, al fine di non indebolire la protezione statistica dei dati.

Si tratta comunque di un percorso complesso. Il D.lgs., se consente indiscutibili miglioramenti sul fronte della fruibilità dei dati, introduce talune complessità gestionali e organizzative già presenti nell'ordinamento europeo, introdotte allo scopo di bilanciare le accresciute possibilità di utilizzo dei dati con più stringenti garanzie.

Le nuove norme prevedono infatti una fase nuova di preliminare riconoscimento delle istituzioni di appartenenza dei ricercatori, sia per l'accesso al laboratorio, sia per le richieste di file MFR, sia per la stipula di protocolli di ricerca. Ai fini di tale riconoscimento è richiesto che le istituzioni, oltre ad attestare, come avviene già oggi, la propria attività di ricerca, sottoscrivano, attraverso un soggetto abilitato a rappresentarle all'esterno, un impegno di riservatezza esteso a tutti i ricercatori che hanno accesso ai dati, nel quale siano specificate le condizioni di utilizzo dei dati, gli obblighi dei ricercatori e i provvedimenti previsti in caso di violazione degli impegni assunti, nonché le misure adottate per tutelare la riservatezza dei dati.

Per dare completa attuazione al D.lgs. sarà quindi necessaria una revisione complessiva del percorso autorizzativo per l'accesso ai dati e – di conseguenza – una revisione della modulistica, tanto per gli MFR quanto per il Laboratorio ADELE. Questa rivisitazione avrà un impatto rilevante anche sul fronte informatico a causa delle modifiche che si dovranno apportare alle procedure attuali.

L'esigenza di una stretta collaborazione tra istituti nazionali di statistica e mondo della ricerca è molto avvertita anche a livello internazionale e la Commissione europea ha lavorato e sta lavorando intensamente per promuovere l'aggiornamento dei metodi e delle tecniche a tutela della riservatezza, ma anche per rivedere l'impianto normativo e le regole che in ambito europeo presiedono all'accesso ai microdati, per consentirne una maggiore e migliore fruibilità.

3.2 Laboratorio ADELE: ulteriori innovazioni

Il mondo della ricerca manifesta da tempo l'esigenza di condurre analisi più approfondite sul settore delle imprese, con la richiesta di poter effettuare studi longitudinali e di disporre di basi di dati che integrino informazioni da più fonti.

Per rispondere a questa esigenza, l'Istituto nel 2011 ha iniziato a rendere disponibili presso il Laboratorio ADELE i primi file integrati di microdati di impresa. Come già segnalato, questa tipologia di file si è progressivamente arricchita, ma siamo consapevoli che questa iniziativa non è comunque sufficiente a colmare le esigenze

conoscitive della comunità scientifica, anche perché non è stato sempre possibile proseguire l'aggiornamento temporale dei data set messi a disposizione.

L'ampliarsi dell'interesse per questo tipo di studi ha spinto l'Istat a cercare nuove soluzioni, puntando sulle potenzialità del Laboratorio ADELE – e in prospettiva – dei Laboratori remoti. L'idea cui si sta lavorando è quella di consentire direttamente ai ricercatori di effettuare *link* tra uno o più file di microdati sulle imprese, così da poter loro stessi creare i file su cui condurre analisi longitudinali o anche integrare dati provenienti da più fonti.

Si tratta di un avanzamento potenzialmente di grande impatto per la ricerca economica, tanto più se si pensa che oggi, in mancanza di questa possibilità⁴, alcuni ricercatori conducono questo tipo di analisi su dataset di altri Paesi.

Il progetto, attualmente in fase di sviluppo, prevede che il codice statistico che Istat attribuisce a ciascuna impresa (abituale denominato *codice ASIA*) sia sostituito da un codice criptato generato al momento dell'accesso al Laboratorio da parte dell'utente autorizzato. Il codice criptato è tale per cui non sarà possibile risalire da questo al codice ASIA e viceversa: l'eventuale conoscenza pregressa di uno o più codici ASIA da parte del ricercatore che accede ad ADELE non offrirebbe quindi alcuna informazione aggiuntiva, potenzialmente utile a fini di intrusione.

In più, la criptazione del codice è relativa al singolo ricercatore; qualora un secondo ricercatore volesse effettuare gli stessi *linkage* avrebbe a disposizione codici criptati diversi da quelli ottenuti dal ricercatore precedente, così da scongiurare la possibilità di un qualsiasi tipo di scambio di informazioni tra utilizzatori diversi degli stessi dati.

Il codice ASIA criptato non si configura come un identificativo diretto; può quindi essere messo a disposizione del ricercatore, il quale potrà e così effettuare nel Laboratorio ADELE i *linkage* previsti dal progetto di ricerca per cui è stato autorizzato⁵.

⁴ Come già detto, i dati identificativi non sono disponibili presso il Laboratorio ADELE.

⁵ A rigore anche il *codice ASIA* non è un identificativo diretto, ma per le sue caratteristiche non offre sufficienti garanzie contro il rischio di intrusione. Il codice ASIA viene infatti utilizzato regolarmente per le comunicazioni tra Istat e imprese. Considerato che in nessun momento le imprese sono state informate della necessità di non renderlo noto e considerata anche la sua persistenza nel tempo, il codice ASIA risulta vulnerabile sotto il profilo della riservatezza, pur non essendo un identificativo diretto.

I risultati delle elaborazioni – e l’affidabilità dei *linkage* ottenuti – restano sotto la responsabilità del singolo ricercatore, mentre – come sempre nel Laboratorio ADELE – resta sotto la responsabilità dell’istituto il controllo finale dell’output ai fini della tutela della riservatezza, prima del suo rilascio.

Gli esiti di questa nuova offerta andranno attentamente monitorati sia dal punto di vista della soddisfazione e dell’utilità per i ricercatori, sia da quello dell’adeguatezza delle risorse e dell’impianto tecnologico messi in campo da Istat.

L’innovazione verrà dunque implementata in modo graduale, a partire da alcune selezionate indagini, preferibilmente svolte con cadenza annuale.

Per poter procedere in questa direzione bisognerà comunque attendere che si completi l’iter di approvazione delle Linee guida per l’accesso ai microdati.

Va ricordato che, negli ultimi anni, l’iniziativa del legislatore di escludere dalla normativa sulla privacy i dati relativi a persone giuridiche, enti o associazioni (art.40 del d.l. n. 201/2011 – convertito con la L. 214/2011) ha comportato per l’Istat una forte criticità, in quanto ha lasciato senza un sicuro riferimento normativo la gestione dell’accesso ai microdati di impresa e delle istituzioni pubbliche e non profit.

Il D.lgs sana questo aspetto, stabilendo che la disciplina per l’accesso ai dati relativi alle persone fisiche (dati personali) si applica ugualmente alle persone giuridiche (imprese e istituzioni, enti o associazioni).

Perché gli avanzamenti appena prospettati possano effettivamente essere messi a disposizione degli utenti, è necessario il parere del Garante, passaggio indispensabile per dare fondamento giuridico a tutte le nuove iniziative in materia di accesso ai microdati di impresa (per il testo completo delle Linee guida si rimanda all’Allegato 5).

3.3 L’informazione sui microdati

Un altro aspetto di rilievo per favorire l’uso dei microdati riguarda l’accesso all’informazione sui microdati stessi.

Su questo tema si è soffermato anche Eurostat nella raccomandazione fatta a conclusione della già citata *peer review*, raccomandazione con la quale l’istituto di statistica europeo invitava l’Istat a migliorare l’informazione resa agli utenti in tema di disponibilità di microdati. In risposta a questa sollecitazione, l’istituto si era impegnato a creare un sito internet specificamente dedicato ai dati elementari.

Tuttavia, alla luce delle nuove strategie dell'Istat in tema di sviluppo dell'informazione statistica e di comunicazione, questa linea non è più adeguata né coerente con l'obiettivo di evitare la creazione di una pluralità di siti *ad hoc*. L'Istituto sta puntando infatti sull'arricchimento del sito istituzionale attraverso l'attenta progettazione di nuove aree al suo interno.

La pagina dedicata ai microdati è stata in effetti già notevolmente ampliata e oggi guida assai meglio gli utenti nel conoscere e reperire le informazioni disponibili.

Le innovazioni più importanti, scaturite da un costante rapporto con gli utenti, hanno riguardato in particolare il Laboratorio ADELE e la possibilità che oggi viene offerta ai ricercatori di consultare liberamente e in dettaglio, attraverso le pagine del sito, tutti i metadati relativi a ciascuna delle indagini dell'Istituto. In questo modo, gli studiosi – ancor prima di chiedere, e ottenere, l'accesso al Laboratorio – sono messi in condizione di conoscere le informazioni su cui potranno effettivamente contare (quali le indagini, quali le tecniche di rilevazione, quali le variabili, quali i tracciati record, eccetera).

Un'apposita sezione del sito guida inoltre i ricercatori nella predisposizione *online* della proposta di progetto di ricerca e nella compilazione della domanda di accesso: un'innovazione importante che ha consentito di ridurre drasticamente i tempi di risposta agli utenti.

Infine, nell'ambito dei lavori per la progettazione del nuovo sito, il cui lancio è programmato per l'inizio del prossimo anno, si prevede che la sezione dedicata ai microdati abbia maggiore evidenza in *home page* e sia più facilmente raggiungibile. Inoltre, attraverso una funzione di ricerca avanzata – basata sia sull'argomento sia sulla data di rilascio – sarà possibile per gli utenti reperire agevolmente le diverse tipologie di file (file standard, file MFR, file a uso pubblico).

L'ipotesi di sviluppare un sito dedicato potrà essere ripresa più avanti, qualora altri enti, anche stimolati dai nuovi dispositivi di legge, vorranno mettere a disposizione i propri microdati. In questa eventualità, l'opportunità di creare un unico punto di accesso, multilogo, che guidi l'utente all'interno di questa offerta più vasta, potrebbe essere riconsiderata.