



**Conservazione e accessibilità dei microdati prodotti dall'Istat e
cenni all'ambito del Sistan**

**Audizione del Presidente dell'Istat
Giorgio Alleva**

**Commissione per la garanzia della qualità dell'informazione statistica
Roma, 12 febbraio 2016**

Indice

	<i>pag.</i>
1. Introduzione	5
2. La conservazione dei microdati delle indagini dell'Istat	6
3. Lo scambio di microdati in ambito Sistan	7
4. La gestione della diffusione, della comunicazione e dell'accesso ai microdati agli utenti non Sistan	8
5. Criticità e iniziative in corso per il miglioramento dell'offerta	11
6. Il processo di modernizzazione della statistica ufficiale e la gestione dei microdati	14

1. Introduzione

Vorrei innanzitutto ringraziare codesta Commissione per l'opportunità che offre all'Istituto di presentare nel dettaglio il nostro impegno sul tema dei microdati. La Cogis, come del resto l'ESGAB (European Statistical Governance Advisory Board) a livello europeo, svolge un ruolo rilevante di monitoraggio e valorizzazione della statistica ufficiale, ruolo di cui noi riconosciamo pienamente il valore di confronto e stimolo per il nostro operato quotidiano. Desidero quindi formulare l'auspicio di una sempre maggiore collaborazione e di un reciproco scambio di conoscenze, con l'obiettivo comune di migliorare sempre più il servizio che l'Istituto e la statistica ufficiale svolgono per il Paese in questa fase che vedrà il riordino del Sistema statistico nazionale e la revisione della programmazione della statistica ufficiale.

In questa audizione illustrerò le diverse fasi della gestione dei dati elementari (microdati) da parte dell'Istat soffermandomi sui risultati, le criticità, le azioni di miglioramento e i prossimi sviluppi anche alla luce delle più recenti evoluzioni tecnologiche, normative e del processo di modernizzazione dell'Istituto.

Prima di entrare nel vivo della trattazione vorrei mettere in luce come la politica di gestione dei microdati da parte dell'Istituto debba contemperare da un lato le esigenze degli utilizzatori – in primis della comunità scientifica – e dei produttori di statistiche, dall'altro quelle di protezione dei dati personali contenute nella cosiddetta normativa sulla privacy, che dal 1996 è entrata a far parte del nostro ordinamento ed è oggi disciplinata dal Codice in materia di protezione dei dati personali (D.lgs. 196/2003).

In quanto fortemente garantista, tale normativa avrebbe potuto rappresentare un cospicuo ostacolo alle possibilità di utilizzo dei dati elementari per scopi statistici e di ricerca scientifica. Per questa ragione,

secondo quanto previsto dalla normativa stessa, nel 2002 è stata delineata una disciplina speciale contenuta nel *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 (Sistan)*, che deroga alle disposizioni generali, in particolare in materia di diritti dell'interessato (informativa e accesso alle informazioni che lo riguardano), conservazione dei dati nel tempo, ecc. e norma anche la comunicazione di dati elementari a soggetti che non fanno parte del Sistan.

2. La conservazione dei microdati delle indagini dell'Istat

I microdati validati prodotti dall'Istat sono archiviati centralmente nell'ARchivio dei MIcroDATi di Istituto, denominato ARMIDA, con il duplice scopo di conservare i microdati e i metadati delle rilevazioni condotte e di favorire il riutilizzo degli stessi per finalità statistiche e di ricerca da parte di utenti esterni e interni.

La procedura adottata per il popolamento di ARMIDA prevede che i settori di produzione dell'Istat, prima di diffondere i dati relativi alle rilevazioni (attraverso comunicati stampa, tavole di dati, volumi, ecc.), procedano sistematicamente all'archiviazione in ARMIDA dei relativi microdati, documentandoli secondo standard predefiniti. L'archivio raccoglie e documenta i microdati relativi a 250 indagini (di cui 95 cessate o sospese) a partire dal 1969 per un totale di circa 12mila file, dei quali 13 derivanti dall'integrazione di più fonti. Quindi, ogni anno vengono archiviati i nuovi dati relativi a tutte le rilevazioni condotte dall'Istituto (circa 160 indagini) alcune delle quali hanno cadenza infrannuale. ARMIDA è un sistema proprietario sviluppato dall'Istituto (basato su software proprietari) e i file di microdati sono memorizzati con codifica ASCII con struttura sequenziale con campi a lunghezza fissa leggibili dagli utenti anche tramite strumenti non proprietari.

Ai fini della conservazione dei dati e in ottemperanza alla normativa vigente, in fase di archiviazione alle variabili è associato un *flag riservatezza* che

identifica diverse tipologie di variabili e i possibili utilizzatori (*codice "identificativo diretto", variabile validata e rilasciabile, variabile sensibile, variabile giudiziaria, variabile non validata oppure variabile di lavoro/controllo*). Inoltre, il sistema separa i dati presenti in base alla loro natura, conservandoli in archivi separati, secondo le modalità previste dalle norme vigenti, laddove ne sia consentita la conservazione (art. 11 del Codice di deontologia e commi 6 e 7 dell'art. 22 del D.lgs. 196/2003). Infine in ARMIDA sono archiviati, in ambienti distinti e dedicati, anche le tipologie di file resi anonimi prodotte dall'Istat per gli utenti esterni (file standard, file per la ricerca, mlcro.STAT).

Dal punto di vista dell'accessibilità, ARMIDA rappresenta la fonte di microdati che consente all'Istituto di rispondere: i) alle richieste degli utenti esterni; ii) alle richieste che pervengono all'Istat da parte di enti facenti parte del Sistan; iii) alle richieste dei ricercatori Istat per il raggiungimento di fini istituzionali.

3. Lo scambio di microdati in ambito Sistan

In base al già citato Codice di deontologia, l'Istat può comunicare ai soggetti Sistan i dati individuali archiviati in ARMIDA senza particolari limitazioni (ovvero anche dati identificativi, se indispensabili al raggiungimento degli scopi del progetto e dati sensibili e giudiziari, se previsto nel Programma statistico nazionale). La procedura per il rilascio dei file agli enti Sistan è molto semplice e avviene attraverso il Contact centre dell'Istat. Nel 2015, le richieste di microdati da parte di enti Sistan all'Istat sono state 370, circa 30 in meno rispetto all'anno precedente ma cinquanta in più rispetto al 2013. La maggioranza dei file sono stati trasmessi agli Uffici di Statistica di Regioni, Province e Comuni.

La circolazione dei microdati in ambito Sistan non si limita però ad un flusso univoco dall'Istat verso l'esterno ma permea l'intero sistema.

L'Istat conduce annualmente la "Rilevazione sugli elementi identificativi, risorse e attività degli uffici di Statistica del Sistan" con specifiche domande sullo scambio dei microdati rivolte agli uffici di statistica dei 653 maggiori enti

escluso l'Istat medesimo¹. I dati relativi al 2014 mostrano che lo scambio di microdati all'interno del sistema è un fenomeno per lo più circoscritto a pochi grandi enti ma in costante aumento. Il 12,6% degli uffici (82 su 653) fornisce dati in media più di 8 volte l'anno mentre il 20% degli uffici (130) richiede i dati in media quasi 5 volte l'anno. Lo scambio ha registrato negli ultimi anni (tra il 2011 e il 2014) una crescita: gli uffici fornitori sono passati dall'8,9 al 12,6% mentre quelli richiedenti dal 14 al 20%. Per quanto garantito dalle norme vigenti, questo flusso non è scevro di difficoltà: ad esempio, il 70% degli uffici che ha richiesto microdati ha segnalato ritardi o tempi lunghi di attesa. Tra gli uffici che non richiedono microdati, una quota consistente ritiene di non averne necessità mentre circa un terzo non ne ha mai valutato la possibilità. Quasi il 13% non possiede le risorse umane per poter trattare i microdati e poco meno del 6% non è a conoscenza della possibilità di interscambio dei microdati.

Infine, il Programma statistico nazionale (Psn) consente di conoscere la tipologia dei flussi di dati elementari all'interno e all'esterno dei confini del Sistema previsti per il 2017. Su 702 lavori inseriti nella programmazione triennale 2017-2019 (in corso di approvazione), 135 (19%) prevedono il rilascio di dati elementari.

4. La gestione della diffusione, della comunicazione e dell'accesso ai microdati agli utenti non Sistan

Per gli utenti non Sistan, l'Istituto ha messo a punto modalità di fruizione dei microdati differenziate per soddisfare diverse tipologie di richieste.

Per quanto riguarda la comunicazione, ovvero la messa a disposizione dei microdati a uno o più soggetti identificati, l'Istat rilascia attraverso il suo Contact centre due tipologie di prodotti informativi: i file per la ricerca (MFR)

¹ Uffici di statistica delle seguenti tipologie istituzionali: Presidenza del consiglio dei Ministri e Ministeri, Prefetture, Camere di commercio, Regioni e Province Autonome, Province, Comuni con oltre 30 mila abitanti, Altre amministrazioni locali e centrali, Altri soggetti pubblici e privati

e i file standard, ora in via di sostituzione con gli MFR. Le due tipologie contengono dati individuali, privi di ogni riferimento che ne permetta il collegamento diretto con gli interessati, e opportunamente trattati per ridurre il rischio di identificazione. La scelta delle indagini per cui realizzare questo tipo di file è caduta su quelle per le quali negli ultimi anni si è concentrata una maggiore domanda da parte della comunità scientifica e/o si sono registrate più richieste di elaborazioni al Contact centre dell'Istituto.

Entrambi i tipi di file sono rilasciati sulla base di uno specifico progetto, i file per la ricerca inoltre sono destinati esclusivamente a ricercatori, purché appartenenti a università o altri istituti di ricerca come previsto dal Codice di deontologia.

L'accesso alla totalità del patrimonio di microdati archiviati in ARMIDA, senza riduzioni né alterazioni del loro contenuto informativo (ad eccezione dei dati identificativi, sensibili e giudiziari), è reso possibile tramite il Laboratorio ADELE (art. 7 comma 2 del Codice di deontologia), situato dal 1998 presso la sede centrale dell'Istat e dal 2011 presso tutte le sedi regionali, ad eccezione di Bolzano. Tengo a sottolineare che le condizioni di utilizzo del Laboratorio, i requisiti e le modalità di accesso, e le regole di rilascio dell'output sono condivise nelle linee essenziali tra i Paesi europei e incluse in un processo di armonizzazione a livello internazionale. Il sito dell'Istituto offre tutte le informazioni necessarie e di recente sono state effettuate modifiche volte a semplificare l'accesso al laboratorio (attraverso una procedura on line guidata) e a rendere chiara l'offerta di informazione disponibile attraverso una "vista" diretta sui metadati di ARMIDA.

Inoltre, è prevista la possibilità di trasmettere dati individuali a organismi aventi finalità di ricerca e non appartenenti al Sistema statistico nazionale nell'ambito di specifici protocolli di ricerca, finalizzati anche al perseguimento dei compiti istituzionali dell'ente Sistan titolare dei dati (art.7 comma 3 del Codice di deontologia). Questa modalità consente di lavorare sui microdati senza dover accedere al laboratorio ADELE. Si basa sui progetti congiunti, finalizzati anche al perseguimento di compiti istituzionali del titolare del

trattamento che ha originato i dati e non consente, come invece accade nel laboratorio Adele, l'accesso alla comunità scientifica nel suo complesso ma solamente ad alcuni ricercatori delle università o istituti di ricerca firmatari dei singoli protocolli ed esclusivamente per i fini in essi contenuti.

Dal 2013, per ampliare il più possibile la platea di potenziali utilizzatori del patrimonio di microdati dell'Istat nel rispetto della normativa vigente, abbiamo introdotto una nuova forma di diffusione che prevede la possibilità per qualsiasi utente di scaricare liberamente e gratuitamente dal sito dell'Istituto, previa accettazione delle condizioni di utilizzo e registrazione, dei File ad Uso Pubblico denominati mlcro.STAT. Questi sono sviluppati a partire dai file per la ricerca, cui vengono applicate tecniche di protezione della riservatezza che implicano una riduzione del contenuto informativo maggiore rispetto ai file per la ricerca.

Nell'ambito della produzione di file ad uso pubblico, l'Italia si pone all'avanguardia nel panorama europeo. Infatti, solo un numero ridotto di paesi offre file ad uso pubblico, e questi vengono essenzialmente prodotti a scopi didattici, dispongono di una documentazione parziale e non assicurano un'affidabilità delle stime tale da consentirne l'uso anche da parte dei ricercatori.

Per chiudere questa rapida panoramica propongo qualche dato sintetico sull'offerta e la domanda di file di microdati.

Nei primi mesi di quest'anno, le indagini per le quali saranno correntemente disponibili sia file per la ricerca sia file mlcro.STAT arriveranno a 12, grazie all'aggiunta dell'Indagine sulle Spese delle famiglie e di quella su Viaggi e vacanze, appena diffuso. La domanda d'altro canto è anch'essa in aumento, in particolare per i file mlcro.STAT, verosimilmente grazie alla facilità di accesso: nel 2015 le richieste di file mlcro.STAT sono state 2.986, quasi il 77% in più rispetto al 2014, mentre quelle dei file per la ricerca sono passate da 70 a 74 per un totale di 370 dataset forniti. Sono in diminuzione invece le richieste di file standard, anche in ragione della loro progressiva sostituzione con gli MFR (da 555 a 342 con un calo pari al 38%) che, comunque

corrispondono a oltre 2.500 file forniti. Tra gli utenti dei file mlcro.STAT, i ricercatori sono poco meno di un quarto (circa 640 individui nel 2015), individui che plausibilmente avrebbero avuto anche i requisiti per chiedere l'accesso ai corrispondenti MFR o ad ADELE. I progetti svolti presso i Laboratori Adele sono invece passati, tra il 2014 e il 2015, da 29 a 31 (erano 18 nel 2013).

L'analisi dell'utenza rende evidente come la facilità di accesso ai microdati costituisca un elemento importante per il sostegno alla ricerca quantitativa, considerazione che è anche alla base delle iniziative intraprese dall'Istituto per avvicinare, anche fisicamente, il laboratorio ADELE ai ricercatori tramite accesso remoto.

Per quanto riguarda invece le indagini più richieste, quasi l'80% delle richieste di file per la ricerca si concentra sull'Indagine sulle forze di lavoro, mentre i file standard si distribuiscono principalmente su 4 grandi indagini: Forze di lavoro, Aspetti della vita quotidiana, Eu-Silc e Consumi delle famiglie. I file mlcro.STAT hanno una composizione più variegata: Forze di lavoro e Aspetti della vita quotidiana sono certamente i più richiesti, ma registrano molte richieste anche le indagini su Innovazione nelle imprese, Incidenti stradali e sui musei (si vedano le tavole allegate).

5. Criticità e iniziative in corso per il miglioramento dell'offerta

Negli ultimi anni, un'importante fonte di criticità è derivata dall'iniziativa del legislatore di escludere dalla normativa sulla privacy i dati relativi a persone giuridiche, enti o associazioni (art.40 del d.l.201/2011 – convertito con la L.214/2011). Ciò ha lasciato la gestione della comunicazione e dell'accesso ai microdati di impresa e delle istituzioni pubbliche e non profit — quindi l'accesso al laboratorio Adele e la stipula di protocolli di ricerca — senza un sicuro riferimento normativo. Con decisione del Consiglio dell'aprile del 2013, l'Istat ha previsto, come soluzione di carattere transitorio, l'utilizzo delle fattispecie individuate dal già citato articolo 7 del Codice di deontologia anche ai dati non regolati dal Codice e riguardanti, quindi, persone giuridiche,

enti e associazioni.

Tuttavia, il perdurare di questa situazione sta mettendo l'Istituto in una condizione di difficoltà, ad esempio rispetto alle proposte di collaborazione provenienti da enti quali la Banca d'Italia e prestigiose università, nonché di inadeguatezza rispetto a principi di ordine generale stabiliti dalla normativa europea (art.23 del Regolamento Ue n.223/2009 del Parlamento europeo e del Consiglio) sull'accesso a dati riservati per fini scientifici.

Alcune criticità sono state rilevate anche nell'ambito della *peer review* condotta da Eurostat presso il nostro Istituto nel corso del 2015 che si è conclusa con una serie di raccomandazioni per il miglioramento della *compliance* dell'Istat al Codice delle statistiche europee. Tra le raccomandazioni, una riguarda l'accessibilità ai microdati: all'Istat è chiesto di prevedere l'accesso ai microdati da remoto e di migliorare l'informazione offerta agli utenti sulla disponibilità dei microdati stessi in ottemperanza al Regolamento Ue n. 557/2013. L'Istituto si è quindi impegnato, attraverso le *Azioni di miglioramento* trasmesse a Eurostat, a creare un sito internet specificamente dedicato ai microdati e soprattutto di lavorare alla predisposizione di laboratori virtuali. Per quanto attiene all'aspetto comunicazione, l'Istituto ha già provveduto a pubblicare sul sito internet una pagina specificatamente dedicata ai microdati accessibili e alla loro documentazione. La realizzazione dei laboratori virtuali, invece, richiede un adeguamento della normativa nazionale.

Un altro aspetto della normativa che ha dimostrato dei limiti riguarda il fatto che l'accesso ai dati elementari a fini di ricerca a soggetti non Sistan sia consentita solo a ricercatori appartenenti ad università e a istituti o enti di ricerca, condizione che talora è apparsa restrittiva portando ad escludere ad esempio gli uffici studi di banche, di organizzazioni di categoria ecc..

Nei mesi scorsi, nell'ambito di un Tavolo tecnico, l'Istat ha avviato un confronto con l'Autorità Garante per la protezione dei dati personali (d'ora in poi Garante) sulle problematiche connesse all'utilizzo per finalità di ricerca dei dati elementari raccolti in ambito Sistan. Il Tavolo ha predisposto una

proposta di modifica normativa, il cui iter di approvazione è in corso di perfezionamento, che delinea nuove regole che consentono di superare molte delle difficoltà descritte. In particolare, si prevede: una più ampia definizione di “ente di ricerca” che comprenda anche istituzioni pubbliche e private o loro specifiche strutture di ricerca; l’istituzione di laboratori virtuali per l’analisi dei dati, accessibili anche da remoto; la previsione di misure adeguate per prevenire e sanzionare eventuali violazioni degli impegni alla tutela della riservatezza assunti dai ricercatori. Infine, si specifica che le regole delineate nel testo saranno da applicarsi anche alle persone giuridiche, enti o associazioni.

Le disposizioni proposte si limitano a dettare norme di principio e indirizzi generali, demandando al Comstat il compito di elaborare apposite linee guida per una puntuale attuazione di una disciplina che si connota per un elevato tecnicismo.

Un ultimo aspetto, ma non per questo meno importante, che desidero richiamare è l’emergere sempre più forte nell’ambito della ricerca dell’esigenza di basi di dati che integrino informazioni provenienti da fonti diverse. Per cercare di venire incontro a queste esigenze, a partire dal 2011 presso il laboratorio Adele abbiamo iniziato a rendere disponibili i primi file di microdati integrati di impresa, risultato di protocolli di ricerca, di progetti internazionali (ad esempio il progetto della BCE CompNet) o di attività di analisi svolte dall’Istituto. Successivamente l’offerta si è arricchita di nuovi file, ma siamo consapevoli che questa iniziativa non è sufficiente a colmare le esigenze conoscitive della comunità scientifica sia sotto il profilo della varietà delle informazioni fornite, sia perché in molti casi non è proseguito l’aggiornamento temporale dei data set messi a disposizione. Tuttavia, la messa a regime di alcuni lavori inseriti nel Psn, quali in particolare la costruzione dell’Archivio di tipo Leed (linked employer employee database) denominato Asia occupazione e la realizzazione della base dati micro TEC-FrameSBS, che integra le informazioni presenti nell’archivio delle imprese esportatrici (TEC) con le principali variabili economiche ora disponibili per

tutte le imprese dell'industria e dei servizi, rappresentano un esempio significativo di come la produzione di basi dati integrate debba entrare a far parte dei processi stessi di produzione statistica.

Tra le ragioni che impediscono lo sviluppo di un maggior numero di basi dati integrate occorre menzionare innanzitutto problemi di natura metodologica, inerenti ad esempio all'integrazione di dati provenienti da diverse indagini campionarie (per evidenti problemi di rappresentatività) o alla verifica della qualità dei dati nel caso di fonti esterne. Inoltre, la possibilità di integrazione tra fonti è limitata dalla normativa in materia di tutela della privacy che impone seri vincoli alla capacità dell'Istituto di produrre ancor prima che di rendere accessibili e di comunicare dati integrati.

6. Il processo di modernizzazione della statistica ufficiale e la gestione dei microdati

L'integrazione delle fonti provenienti da diverse istituzioni e il conseguente sfruttamento a fini statistici dell'informazione di fatto già disponibile rappresenta una delle linee strategiche del mio mandato e del processo di modernizzazione dell'Istat che, proprio in questi giorni, sta vedendo la luce. Il progetto che stiamo perseguendo 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). La modernizzazione del processo di produzione delle statistiche ufficiali richiederà l'armonizzazione di concetti e definizioni, essenziale per procedere all'integrazione di dati provenienti da fonti diverse e permetterà di superare, in alcuni casi, la frammentazione ancora esistente fra diverse indagini campionarie. L'integrazione tra tutte le fonti permetterà di ottenere: dati e indicatori per differenti dimensioni e sotto-popolazioni, senza costi aggiuntivi e riducendo la pressione statistica sui rispondenti; quadri informativi multidimensionali e longitudinali con informazioni più dettagliate;

modelli interpretativi più efficaci grazie alla maggiore disponibilità di informazione sulla interconnessione dei fenomeni (lavoro-istruzione; redditi-consumi-ricchezza; ecc.); un sistema di indagini sociali integrate con il censimento della popolazione; indicatori anticipatori del ciclo economico e della congiuntura attraverso lo sfruttamento delle nuove fonti di dati (in particolare attraverso l'utilizzo dei Big Data, di cui accenneremo brevemente più avanti).

La buona riuscita di questo programma, che prende a modello gli Istituti Nazionali di Statistica dei paesi del Nord Europa ed è in linea con le raccomandazioni fatte da codesta Commissione nell'ambito dell'ultimo parere al "Psn 2014-2016. Aggiornamento 2016", ha due presupposti fondamentali. Da un lato, prevede il superamento delle logiche proprietarie dei dati a favore di un progetto comune, ascrivibile a tutte le istituzioni coinvolte; in quest'ottica desidero ricordare la recente firma di un accordo di collaborazione tra Istat, Inps, Inail e il Ministero del lavoro e delle politiche sociali per lo scambio automatizzato delle informazioni contenute nei rispettivi archivi amministrativi e statistici al fine di migliorare la qualità delle statistiche in tema di occupazione. Dall'altro lato, interroga fortemente l'adequazione delle norme in materia di privacy a supportare i nuovi assetti produttivi, all'interno di una più generale riflessione sull'ordinamento statistico nazionale. Le regole sulla raccolta e il trattamento dei dati devono poter continuare ad evolvere verso un modello in grado di consolidare la fiducia degli individui contenendo il più possibile il rischio di un impiego pregiudizievole delle informazioni (con negativi effetti sulla propensione del singolo a "cedere" i propri dati); al tempo stesso, devono favorire lo sviluppo di nuove forme di acquisizione e gestione delle informazioni (in particolare facilitando l'accesso alle nuove fonti di dati e la loro integrazione) al fine di accrescere l'efficienza e migliorare la qualità delle statistiche prodotte e diffuse, così come richiesto, da ultimo, dal Regolamento UE 759/2015 del Parlamento europeo e del Consiglio, del 29 aprile 2015, che modifica il Regolamento UE 223/2009, relativo alle statistiche europee. Determinanti, in

questa fase, potranno risultare gli orientamenti del Garante della privacy per l'individuazione di adeguate soluzioni applicative o, eventualmente, per elaborare correttivi di carattere normativo da proporre nelle competenti sedi istituzionali.

Mi preme sottolineare che alcuni recenti pareri del Garante hanno evidenziato una serie di difficoltà che occorre affrontare e superare al fine di evitare che le regole attualmente vigenti diventino un ostacolo per l'efficace e compiuta realizzazione del progetto. In occasione dell'esame del "Psn 2014-2016. Aggiornamento 2016" sono stati evidenziati profili problematici sulla possibilità di integrare in sistemi informativi più fonti amministrative di dati individuali in quanto questa pratica, seppur riducendo costi e fastidio statistico, espone gli individui, usando le parole del Garante, "ad un costante incremento del rischio di reidentificazione in relazione a dati in principio aggregati o anonimi". Il Garante esprime preoccupazione anche sul fatto che, per garantire la possibilità di ampliare nel tempo i sistemi informativi, in essi vengano conservati i dati identificativi. Questo parere è espresso in particolare per quanto riguarda il nostro progetto SIM, un repository di dati costruito mediante l'integrazione di microdati di fonte amministrativa e statistica con l'obiettivo di supportare i processi di produzione dell'Istat in un'ottica di standardizzazione e centralizzazione di alcune funzioni produttive trasversali (acquisizione, archiviazione...). Il Garante in questo ambito fa anche specifico riferimento all'utilizzo dei Big data e richiama a tal proposito un parere, altrettanto critico, dell'organo consultivo indipendente dell'Unione Europea in materia di protezione dei dati personali. Inoltre, nel citato parere al Psn e in un ulteriore parere in materia di censimento della popolazione e di Archivio nazionale dei numeri civici e delle strade urbane, il Garante ha espresso un parere negativo sulla possibilità di procedere alla revisione post censuaria delle anagrafi attraverso i dati raccolti dal censimento nella nuova forma che quest'ultimo sta assumendo (censimento permanente).

In questo quadro occorre ricordare che la normativa generale sulla privacy è in corso di revisione anche a livello europeo, soprattutto per cercare di adeguare la normativa alle sfide imposte dalla tecnologia. Con specifico riferimento ai problemi della statistica e della ricerca ricordo che tra gli istituti di statistica europei è costantemente vivo il dibattito sulle modalità di scambio e fruizione di dati elementari ed è attiva una task force che sta definendo i “Core principles for the exchange of identifiable micro-data” in materia di dati di impresa. Già da ora, tuttavia, uno dei problemi che si pone con maggior forza è senza dubbio quello dell’uso a fini statistici dei Big data.