Università degli Studi di Padova

Dipartimento di Diritto Pubblico, Internazionale e Comunitario

Dipartimento di Scienze Statistiche

Corso di Laurea in Diritto e Tecnologia

a.a. 2022/2023

PRIVACY E RICERCA SCIENTIFICA: UN PERCORSO VERSO L'EQUILIBRIO NELL'UTILIZZO RESPONSABILE DEI MICRODATI

Relatore: Prof. Tommaso Di Fonzo

Studente: Loris Palmarin

Sommario

Introduzione

La rapidità con cui evolvono le nuove tecnologie e l'ingente quantità di dati generati quotidianamente dal naturale funzionamento della società hanno posto nuovi interrogativi riguardo alla protezione del diritto dell'individuo a godere della propria privacy. In particolare, questo legittimo diritto individuale rischia di scontrarsi con quello alla ricerca scientifica. Da un lato il rispetto della privacy della persona è un principio fondamentale, dall'altro la ricerca scientifica riveste un ruolo cruciale nello sviluppo della conoscenza e nell'allargamento dei confini della scienza.

La presente tesi si propone di esaminare la necessità di trovare un equilibrio tra questi due diritti, concentrandosi su metodi di accesso e tecniche di utilizzo dei microdati.

Nel capitolo 1, si analizzano le basi giuridiche sia del diritto alla privacy che di quello alla ricerca scientifica. Nel primo caso, si mettono in luce l'importanza della tutela dei dati personali e della riservatezza delle informazioni, nel secondo si evidenzia il ruolo della ricerca nel progresso sociale e scientifico. Inoltre, si analizza la necessità di bilanciare questi due diritti, riconoscendo le sfide che sorgono nel contesto attuale.

Nel capitolo 2, si approfondisce l'utilizzo dei microdati per la ricerca scientifica e si esplorano le diverse fonti da cui possono essere tratti. Un'attenzione particolare verrà dedicata al riutilizzo dei dati per fini di ricerca e alle tecniche di pseudonimizzazione per garantire la riservatezza dei dati coinvolti. Il capitolo 3 si concentra nell'analisi di alcuni casi studio sull'utilizzo dei microdati: si mettono a confronto l'esperienza francese, che si distingue per un sistema in cui un gestore centralizzato coordina la produzione dei dati da parte di diversi produttori, e l'esperienza danese, che sfrutta il potenziale dei registri amministrativi per fini di ricerca.

Nelle conclusioni, viene confrontata la situazione italiana con quella europea, evidenziando le sfide comuni e le differenze emerse in questo contesto. Si delinea anche una visione dei possibili sviluppi futuri, con particolare attenzione alle opportunità offerte dalla tecnologia blockchain per garantire la tutela della privacy.

Capitolo I

Il diritto alla privacy e il diritto alla ricerca scientifica

Sommario: 1.1. Nozione giuridica e aspetti etici relativi alla privacy – 1.2. Nozione giuridica e aspetti etici relativi alla ricerca scientifica – 1.3. Necessità di bilanciamento tra privacy e ricerca scientifica

1.1. Nozione giuridica e aspetti etici relativi alla privacy

Sin dalla sua prima definizione come "diritto ad essere lasciati soli"[1], la privacy si manifesta come un tema fondamentale nella società moderna, tanto da guadagnare il titolo di diritto umano. Con l'avvento di Internet e delle tecnologie digitali, i dati personali sono diventati sempre più vulnerabili ad abusi e violazioni della privacy. In Europa, l'importanza della privacy è stata riconosciuta a livello normativo attraverso la Carta dei diritti fondamentali dell'Unione europea e il Regolamento generale sulla protezione dei dati (GDPR), che è entrato in vigore nel 2018. Tuttavia, nonostante questa prima barriera giuridica, la protezione della privacy continua a essere una sfida importante per le società europee, che cercano tuttora di definire e tutelare un concetto poliedrico e mutevole.

La parola "privacy" viene utilizzata per la prima volta da due giovani avvocati americani, Warren e Brandeis, nel saggio *The right to privacy*, pubblicato nel 1890. Essi percepiscono la necessità di proteggere la facoltà di ogni essere umano di essere lasciato solo, di poter "godere del proprio" e decidere di non condividere fatti personali, pensieri ed emozioni. Proprio su questa pubblicazione si baserà il sistema americano per individuare quattro illeciti legati a questo concetto[2]. L'Europa, invece, inizia ad analizzare il diritto alla privacy dopo gli Stati Uniti, sviluppando un tipo diverso di protezione.

Non esiste una definizione ufficiale di privacy: si tratta di un concetto universale, che vuole trasmettere un determinato contenuto, ma fortemente influenzato dagli aspetti socio-culturali ed economici in cui si sviluppa. Daniel Solove, professore di diritto alla Washington University Law School, individua sei categorie per le numerose definizioni del termine, secondo le quali la privacy è (1) il diritto ad essere lasciati soli, (2) l'accesso limitato al sé, (3) segretezza, (4) controllo di informazioni personali, (5) personalità e (6)

[1] Warren and Brandeis, *The Right to Privacy*, 1890.

[2] "*4 categories: intrusion into seclusion, appropriation of name or likeness, public disclosure of private facts* and *placing a person in a false light"*. American Law Institute, *Restatement (2nd) of Torts*, §652.

intimità[3]. Secondo l'autore, questa suddivisione permette di definire il concetto di "privacy" in maniera più flessibile, facilitando così il lavoro dei tribunali nell'affrontare al meglio i problemi emergenti e in costante evoluzione.

Nonostante questa incertezza nella definizione, diversi documenti internazionali riconoscono il diritto alla privacy che, di conseguenza, inizia ad apparire anche nelle norme di recepimento delle legislazioni nazionali. Ai fini di questa tesi, i riferimenti normativi presi in considerazione saranno quelli aventi effetti nel contesto europeo e, nello specifico, quello italiano. L'articolo 12 della Dichiarazione Universale dei Diritti Umani (DUDU), l'articolo 8 della Convenzione Europea dei Diritti dell'Uomo (CEDU) e l'articolo 8 della Carta dei Diritti Fondamentali dell'Unione Europea (o Carta di Nizza) stabiliscono che il diritto alla privacy è un diritto umano fondamentale e che ognuno ha il diritto a goderne.

Firmata nel 1950 dal Consiglio d'Europa, la CEDU è un trattato internazionale che ha come obiettivo quello di tutelare i diritti umani e le libertà fondamentali ("EUR-Lex - eu_human_rights_convention - EN - EUR-Lex - Europa"). Proprio nell'articolo 8 compare un primo tentativo di protezione della privacy, che tutela quattro interessi: la vita privata, la vita familiare, il domicilio e la corrispondenza. Tale articolo è volto alla tutela dell'ingerenza arbitraria nella vita privata e assegna all'autorità pubblica un obbligo negativo (un'astensione), che però assume anche un'accezione positiva nel garantire che i diritti previsti da questo articolo siano rispettati anche tra privati[4].

Il concetto di privacy vede una delle sue prime estensioni a "protezione dei dati personali" a seguito della promulgazione della Carta dei Diritti Fondamentali dell'Unione Europea, nota anche come Carta di Nizza, avvenuta nel 2000. La Carta definisce i diritti civili di cui godono i cittadini europei. L'articolo 7 conferma la visione della CEDU sostenendo che "*ogni persona ha diritto al rispetto della propria vita privata e familiare, del proprio domicilio e delle proprie comunicazioni*".

---

[3] Daniel Solove, *Conceptualizing Privacy*, 90 California Law Review 1087, 2002.

[4] Articolo 8 della Convenzione– Diritto al rispetto della vita privata e familiare "1. Ogni persona ha diritto al rispetto della propria vita privata e familiare, del proprio domicilio e della propria corrispondenza. 2. Non può esservi ingerenza di una autorità pubblica nell'esercizio di tale diritto a meno che tale ingerenza sia prevista dalla legge e costituisca una misura che, in una società democratica, è necessaria alla sicurezza nazionale, alla pubblica sicurezza, al benessere economico del paese, alla difesa dell'ordine e alla prevenzione dei reati, alla protezione della salute o della morale, o alla protezione dei diritti e delle libertà altrui."

A differenza degli anni Cinquanta, però, il concetto di privacy non è più attuale e "nuovo", ed ha bisogno di essere rimodellato adattandolo all'evolversi delle tecnologie e alle sfide che queste propongono. Avviene pertanto una ri-contestualizzazione che non prevede l'eliminazione della sfera privata e confidenziale ma l'integrazione di questa ad un nuovo mondo caratterizzato da social media e connessioni onnipervasive. L'articolo 8, infatti, affronta questa necessità emergente affermando che "*ogni persona ha diritto alla protezione dei dati di carattere personale che la riguardano*". Il secondo comma prosegue stabilendo che "*tali dati devono essere trattati secondo il principio di lealtà, per finalità determinate e in base al consenso della persona interessata o a un altro fondamento legittimo previsto dalla legge*"[5]. Grado di consapevolezza e libertà dell'utente nella manifestazione del consenso assumono un ruolo fondamentale in una società che passa dalla ricerca dei dati alla necessaria selezione degli stessi a causa dell'iperproduzione che la caratterizza.

Un cambio di ritmo notevole si registra nel 2016 con la promulgazione del Regolamento Generale sulla Protezione dei Dati (GDPR), anche conosciuto come Regolamento UE 2016/679. Esso disciplina il trattamento dei dati personali relativi alle persone nell'UE, da parte di persone, società ed organizzazioni[6]. Al contrario delle direttive, il regolamento è immediatamente efficace in tutti i paesi aderenti all'Unione Europea. Agli stati membri è stato concesso un periodo di due anni per il recepimento e l'adattamento delle normative nazionali. Il GDPR interviene definendo rigide regole per il trattamento dei dati: si applica a tutti i soggetti che elaborano dati di natura personali di residenti nell'UE, definisce precisi doveri e responsabilità al titolare del trattamento dei dati e infligge sanzioni in caso di violazioni. Nei paragrafi successivi, si tratteranno gli articoli che coinvolgono direttamente la ricerca scientifica.

[5] Carta dei Diritti Fondamentali dell'Unione Europea, 2000.

[6] https://commission.europa.eu/law/law-topic/data-protection/reform/what-does-general-data-protection-regulation-gdpr-govern

1.2. Nozione giuridica ed aspetti etici relativi alla ricerca scientifica

"Le arti e la ricerca scientifica sono libere. La libertà accademica è rispettata"[7]. L'articolo 13 della Carta dei diritti fondamentali dell'Unione Europea è chiaro e garantisce, all'interno dei paesi europei, il diritto insindacabile alla ricerca scientifica.

Per poterne discutere, è necessario definire cosa comprende la ricerca scientifica. È un processo sistematico e metodologico finalizzato alla scoperta, alla verifica e alla validazione di nuove conoscenze e teorie riguardanti un determinato ambito di studio[8]. Per mezzo della ricerca, è possibile allargare i confini della scienza aumentando la comprensione di fenomeni o sviluppando nuove tecnologie in grado di migliorare la vita delle persone.

Il regime giuridico della ricerca scientifica in Europa è complesso e varia da paese a paese. Riconosciuta unitariamente come un'attività di grande importanza per il progresso sociale, gli stati membri si impegnano a sviluppare sistemi di regole e normative per garantire l'etica e la legalità.

La Convenzione 108 di Strasburgo per la protezione delle persone fisiche con riguardo al trattamento automatizzato dei dati personali del 1981, la Carta dei diritti fondamentali dell'Unione Europea e il Regolamento UE 679/2016 (GDPR) sulla protezione dei dati personali costituiscono la cornice degli strumenti giuridicamente vincolanti che racchiude i limiti della ricerca scientifica. L'Unione Europea si impegna, dunque, ad affrontare e trattare gli aspetti etici e legali legati a questa attività.

Il Regolamento Generale sulla Protezione dei Dati (GDPR) dell'UE definisce un sistema di regole per la gestione dei dati personali dei cittadini europei, citando anche quelli che vengono poi utilizzati per la ricerca scientifica. Sin dalle prime considerazioni dell'atto, infatti, è riconosciuto un ruolo fondamentale alla ricerca, che viene identificata come un ottimo strumento per raccogliere conoscenze su determinate condizioni sociali (disoccupazione, livello di istruzione) e, proprio per questo, deve essere facilitata[9].

Il Considerando 159, invece, aiuta a definire il concetto di ricerca scientifica sottolineando il fatto che va interpretato in senso lato includendo, ad esempio, sviluppo

[7] Art.13, Carta dei Diritti Fondamentali dell'Unione Europea.

[8] Paolo Bisogno, *Enciclopedia Italiana Treccani*, IV Appendice, 1981.

[9] Considerando 157, GDPR: "*Al fine di facilitare la ricerca scientifica, i dati personali possono essere trattati per finalità di ricerca scientifica fatte salve condizioni e garanzie adeguate previste dal diritto dell'Unione o degli Stati membri.*"

tecnologico, ricerca fondamentale, applicata, finanziata da privati e svolta nell'interesse pubblico. Proprio questo riferimento normativo ricorda, inoltre, che il Trattato sul Funzionamento dell'Unione Europea (TFUE) si è posto come obiettivo quello di "*rafforzare le sue basi scientifiche e tecnologiche con la realizzazione di uno spazio europeo della ricerca nel quale i ricercatori, le conoscenze scientifiche e le tecnologie circolino liberamente, di favorire lo sviluppo della sua competitività, inclusa quella della sua industria, e di promuovere le azioni di ricerca ritenute necessarie ai sensi di altri capi dei trattati*"[10]. Tutti gli articoli a seguire evidenziano la volontà dell'UE di progredire nell'ambito della ricerca definendo norme comuni ed eliminando ostacoli giuridici e fiscali[11].

L'Unione Europea, infine, favorisce la cooperazione internazionale per la ricerca scientifica garantendo sia la libera circolazione dei ricercatori che quella dei risultati della ricerca. In questo modo, si cerca di garantire risultati di alta qualità, promuovendo una ricerca responsabile e rispettosa dei diritti degli individui coinvolti.

1.3. Necessità di bilanciamento tra privacy e ricerca scientifica

Dopo aver definito due concetti molto attuali, privacy e ricerca scientifica, si è visto come questi possano essere considerati veri e propri diritti tutelati da stringenti regimi normativi. È inevitabile, tuttavia, che la copresenza di diritto alla privacy e diritto alla ricerca scientifica necessiti di un'operazione di confronto tra le normative e di identificazione di un equilibrio che integri e contemperi i due regolamenti ottenendo un risultato che possa tutelare sia i diritti dell'individuo che quelli del ricercatore il quale, come già accennato, agisce negli interessi della scienza.

Il protrarsi di situazioni di incertezza relativamente all'applicabilità della normativa in tema privacy hanno reso quella esistente "precocemente invecchiata" e, di conseguenza, obsoleta. Proprio in questo contesto, interviene il GDPR che segna un punto di svolta e aiuta a creare una cornice giuridica più coerente stabilendo una serie di regole per la gestione e la protezione dei dati personali dei cittadini europei, inclusi i dati utilizzati per la ricerca scientifica.

[10] Art.179, par.1, TFUE.
[11] Art.179, par.2, TFUE.

Questo ragionamento vede il suo principio nel Considerando n.4 che si pone l'obiettivo di adattare la protezione dei dati con lo sviluppo economico e tecnologico, sancendo che: "*Il diritto alla protezione dei dati di carattere personale non è una prerogativa assoluta, ma va considerato alla luce della sua funzione sociale e va contemperato con altri diritti fondamentali, in ossequio al principio di proporzionalità.*"[12]. Quindi, il focus non è solo quello della protezione dei dati ma la necessità di individuare una stabilità tra il rilievo della privacy dell'individuo e l'utilizzo a fini sociali[13]. La ricerca scientifica rientra sicuramente in questi fini.

A proposito della propulsione che vuole essere data allo sviluppo economico e tecnologico, è importante ricordare che l'Unione Europea si è già espressa in materia. Il 25 Aprile 2018, la Commissione Europea ha emanato un comunicato stampa che ha ad oggetto l'impegno della stessa ad incrementare la facilità d'accesso ai dati e la disponibilità degli stessi.

Andrus Ansip, vicepresidente della Commissione e responsabile per il Mercato unico digitale, ha dichiarato che l'intelligenza artificiale e tecnologie simili saranno in grado di aiutare l'Europa a migliorare numerosissimi aspetti sociali, tra cui il risparmio energetico, l'istruzione e l'assistenza sanitaria[14]. L'obiettivo è chiaro: costruire un'economia dei dati, volta alla crescita dei mercati e all'aumento del benessere.

Appurata, quindi, la forte propensione dell'Unione alla creazione di un mercato che abbia come principale creatore di valore il dato, è necessario analizzare come il legislatore europeo si è mosso per bilanciare i diritti dell'individuo con l'auspicato progresso. Il cardine normativo su cui si basa questa tesi è dunque il Regolamento Generale per la Protezione dei Dati (GDPR) che stabilisce una serie di regole per la raccolta, l'elaborazione e la conservazione dei dati utilizzati per la ricerca scientifica, al fine di garantire un approccio etico e rispettoso dei diritti dei soggetti coinvolti.

Il perno della normativa legata alla ricerca scientifica si colloca all'articolo 89, che viene richiamato frequentemente da articoli precedenti che prevedono deroghe ai principi

[12] Considerando n.4, GDPR.

[13] Avv. Silvia Stefanelli, *Ricerca scientifica e privacy. I limiti europei sono sufficienti. Perché l'Italia vuole andare oltre?*, in *Quotidiano Sanità*, 2018.

[14] Comunicato Stampa della Commissione Europea, *Dati nell'UE: forte impegno della Commissione per aumentare la disponibilità dei dati e promuoverne la condivisione nel campo dell'assistenza sanitaria*, Bruxelles, 2018.

generali sanciti dal Regolamento. In tale articolo, viene stabilito che per trattare i dati personali con finalità di archiviazione nel pubblico interesse, di ricerca scientifica o storica o a fini statistici, il titolare del trattamento deve attuare adeguate tutele per i diritti e le libertà degli interessati. Tali garanzie possono includere l'uso di tecniche di pseudonimizzazione o anonimizzazione, nonché l'attuazione di adeguate misure tecniche e organizzative per garantire la sicurezza dei dati.

L'articolo 89 non è importante per le sole deroghe che prevede, ma anche per le finalità che intende perseguire. Alcuni autori riconoscono, infatti, una doppia natura di questa norma: da una parte quella programmatica, volta a diventare una linea guida per i legislatori nazionali ed europei che sono chiamati ad adottare i principi di minimizzazione e di non identificabilità dell'interessato, dall'altra quella precettiva, immediatamente vincolante, che ordina il rispetto di questi principi[15].

È opportuno sottolineare che la maggior parte dei dati a disposizione del ricercatore è già disponibile e rappresenta un'alternativa economicamente vantaggiosa per il lavoro di ricerca. Questa sembra essere la ratio che sottende l'articolo 5, lettera (b), il quale sancisce il principio di limitazione delle finalità ed un'eccezione: i dati personali devono essere raccolti per finalità determinate, esplicite e legittime e il riuso di questi dati a fini di ricerca scientifica o storica o a fini statistici non è considerato incompatibile con le finalità iniziali[16].

Nello stesso articolo, la lettera (e) stabilisce anche una deroga alla limitazione della conservazione dei dati, circoscritta al tempo necessario al conseguimento delle finalità. Viene garantita una maggiore libertà al ricercatore fornendo la possibilità di conservare i dati in una forma che permetta l'identificazione degli interessati anche oltre il raggiungimento dell'obiettivo e nella piena osservazione di tutte le cautele che il regolamento richiede.

Un'ulteriore eccezione viene fissata dall'articolo 17, paragrafo 3. Il GDPR garantisce all'interessato il diritto alla cancellazione dei dati personali che lo riguardano, anche chiamato "diritto all'oblio". Se l'eliminazione dei dati rischia di minare il conseguimento degli obiettivi stabiliti dalla ricerca, è possibile imporre la non applicazione del suddetto

[15] V. Cuffaro, et al. *I dati personali nel diritto europeo*. G. Giappichelli, 2019, 573-578.

[16] S. Ruggieri, *Towards a Digital Ecosystem of Trust: Ethical, Legal and Societal Implications*, in *Opinio Juris In Comparatione*, 1, 2021.

diritto. Sarà necessario quindi tenere conto della portata della ricerca, della dimensione dei campioni analizzati e dell'influenza che hanno i dati del soggetto coinvolto nel perseguimento dell'obiettivo finale.

In sintesi, il GDPR riconosce il valore pubblico della ricerca scientifica e l'importanza che questa ricopre nel progresso economico e tecnologico, ma impone alcune restrizioni sulla raccolta, l'elaborazione e la conservazione dei dati utilizzati appartenenti all'interessato. La normativa definisce ed impone l'individuazione di figure quali il responsabile e il titolare del trattamento, che svolgono un ruolo cruciale nel trattamento, oltre che l'adozione di numerosi requisiti tecnico-organizzativi nell'intero flusso dei dati. Questi temi verranno affrontati nel corso della tesi ponendo particolare attenzione alla categoria del microdato, termine utilizzato dalla statistica per riferirsi a unità elementari di informazione che contengono informazioni su individui, famiglie o imprese[17].

[17] Definizione ISTAT, https://www.istat.it/it/dati-analisi-e-prodotti/microdati

Capitolo II

Utilizzo di microdati per la ricerca scientifica

Sommario: 2.1. Definizione, fonti, utilizzo dei microdati. – 2.2. Dare nuova vita ai dati: il riutilizzo per fini di ricerca – 2.3. Tecniche di pseudonimizzazione dei dati

"*Mentre il singolo individuo è un enigma irrisolvibile, quando è insieme agli altri diviene una certezza matematica. È impossibile, per esempio, predire il modo in cui agirà un uomo, mentre è invece possibile dire con precisione cosa faranno un certo numero di uomini messi insieme. L'individuo varia, ma le percentuali rimangono costanti. Così dicono le statistiche.*" Citando le parole dello storico inglese William Winwood Reade, il celebre investigatore Sherlock Holmes nel romanzo di Arthur Conan Doyle aiuta il lettore a comprendere il valore della statistica, disciplina che serve la ricerca scientifica e permette di estrarre informazione dai dati che, se non aggregati, sono privi di valore.
Questa sezione mira ad introdurre il concetto di microdato definendolo ed elencandone fonti e utilizzi in campo scientifico. Segue un'analisi dei vantaggi del riutilizzo dei dati amministrativi per la ricerca scientifica.

2.1. Definizione, fonti, utilizzo dei microdati

I microdati, nella statistica, sono dati elementari ed individuali. Sono chiamati "micro" perché forniscono informazioni su singoli individui o unità statistiche, come famiglie o imprese, e si contrappongono alla dimensione "macro" che rappresenta un'elaborazione dei dati che si riferisce a campioni o popolazioni. Possono includere informazioni personali, ad esempio l'età, il genere, la professione e altre informazioni socio-demografiche. Possono essere utilizzati per scopi di ricerca scientifica, per esempio per valutare l'impatto di politiche pubbliche o per analizzare abitudini di consumo.
Esistono diverse fonti dalle quali è possibile estrarre i microdati. Il metodo più semplice è quello che ricorre a sondaggi e censimenti, utilizzati per raccogliere informazioni su aspetti quali la salute, l'occupazione, il reddito. Possono essere impiegati per analisi di ricerca o per l'applicazione e la verifica di politiche pubbliche. Un'altra buona fonte è costituita dai registri amministrativi, ovvero raccolte di dati che vengono collezionate da istituzioni governative. Di notevole utilità sono anche i dati delle aziende su prodotti, servizi, clienti e dipendenti. Infine, riferendosi al mondo digitale, dati provenienti da fonti

web, come i social network, permettono di ottenere informazioni dettagliate sui comportamenti e le preferenze degli utenti.

Il Regolamento della Commissione Europea n.557/2013 rappresenta uno strumento normativo fondamentale per la gestione dei microdati derivanti dalle rilevazioni statistiche e mira ad armonizzare le normative degli Stati membri dell'Unione europea. Aiuta, infatti, a definire e categorizzare i microdati, introducendo i "file per uso sicuro" e i "file per uso scientifico"[18]. Una visione coerente che, per quanto possibile, armonizzi le categorie di microdati potrebbe essere la seguente[19]:

- Dati ad uso pubblico (PUF), definiti dati "resi anonimi e predisposti in modo tale che le unità statistiche non possano essere identificate, direttamente o indirettamente, tenuto conto di tutti i pertinenti mezzi che possono essere ragionevolmente utilizzati da un terzo".[20] Ad essi, essendo già anonimizzati, non si applica la normativa sulla privacy.
- Dati riservati destinati a fini scientifici (*confidential data for scientific purposes*). Questi non contengono indicatori che permettano l'identificazione diretta dell'interessato e ad essi afferiscono due classi di microdati:
    - File per uso scientifico (MFR)[21]. Questi sono dati a cui sono stati applicati metodi volti al controllo della garanzia di riservatezza. L'accesso a questi file è garantito per lo sviluppo di un progetto di ricerca ad opera di ricercatori appartenenti ad organizzazioni riconosciute dal Comstat.
    - File per uso sicuro (SUF). Questo sono "*i dati riservati destinati a fini scientifici cui non sono stati applicati ulteriori metodi di controllo della diffusione statistica*"[22].
- Dati con identificativi diretti[23]. Questa categoria di dati è conservata solo dall'ente produttore o dal gestore e fa riferimento alla versione del dato nella sua forma più

[18] Art.2, Regolamento (UE) n.557/2013 della Commissione Europea.
[19] https://www.istat.it/it/dati-analisi-e-prodotti/microdati
[20] Art.19, Regolamento (CE) n.223/2009 del Parlamento Europeo e del Consiglio.
[21] Art.2, Regolamento (UE) n.557/2013 della Commissione Europea.
[22] Art.2, Regolamento (UE) n.557/2013 della Commissione Europea.
[23] U. Trivellato, Data-Driven Policy Impact Evaluation: How Access to Microdata is Transforming Policy Design. Germania, Springer International Publishing, 2018.

grezza, che contiene, cioè, elementi che permettono l'identificazione dell'interessato.

Come già analizzato, il TFUE ha scandito un cambio di ritmo nell'evoluzione dell'accesso ai dati da remoto, stimolando la creazione di uno spazio europeo accessibile da ricercatori autorizzati in cui i microdati superino i confini nazionali e siano disponibili a tutti gli Stati membri. Comincia, quindi, una fase di sperimentazione dei servizi di accesso remoto ai dati.

Un primo modello è quello dell'esecuzione remota. Il ricercatore accreditato trasmette via e-mail un file da elaborare scritto in un linguaggio statistico ad un "safe center", il quale si occupa dell'elaborazione dei dati e ritorna l'output al ricercatore. Questa possibilità garantisce la sicurezza dei dati perché non escono dai centri di elaborazione, ma sono molto limitanti per il ricercatore perché ha poca interazione con i dati. Un caso esemplare di esecuzione remota è quello dello IAB FDZ, un centro di ricerca tedesco specializzato nella fornitura di microdati anonimizzati per scopi di ricerca, principalmente sul mercato del lavoro, che nel 2014, ad esempio, ha elaborato circa 1800 richieste[24].

Un secondo modello è quello dell'accesso decentralizzato ad un centro sicuro. Il ricercatore accede direttamente ai dati in "safe rooms", presso l'ente che li gestisce oppure presso membri della sua rete, che possono essere altri istituti di ricerca o centri universitari. Un esempio, che verrà approfonditamente analizzato in seguito, riguarda l'applicazione di questo metodo nel laboratorio ADELE in Italia, una struttura specializzata nella gestione dei microdati provenienti da indagini statistiche ISTAT.

Il terzo modello riguarda l'accesso ai dati in senso proprio, ovvero tramite l'identificazione di ricercatori accreditati che possono lavorare sui dati direttamente dai propri strumenti, utilizzando dei software che garantiscono sicurezza ed affidabilità. In questo modo, si configura il miglior equilibrio tra la garanzia di sicurezza nella gestione dei dati e la possibilità per il ricercatore di interfacciarsi direttamente con i dati. Una buona implementazione di questa tipologia si è ottenuta in Danimarca nella

[24] D. Muller & J. Moller, *Giving the International Scientific Community Access to German Labor Market Data: A Success Story* in *Data-Driven Policy Impact Evaluation: How Access to Microdata is Transforming Policy Design*, Germania, Springer International Publishing, 2018.

Statistikbanken, agenzia statistica nazionale responsabile della diffusione di statistiche ufficiali sul paese[25].

2.2. Dare nuova vita ai dati: il riutilizzo per fini di ricerca

La capacità di estrarre valore dei dati è legata alla capacità di integrare dati che provengono da fonti differenti. Le fonti non devono essere sempre e necessariamente "nuove", ma possono riferirsi a set di dati già esistenti, che risultano più economici e immediati.

Fornire la possibilità di accedere ai microdati è un passo fondamentale nello sviluppo della ricerca scientifica. Questo, però, richiede la possibilità per i ricercatori di riutilizzare dati già esistenti, provenienti da altre ricerche scientifiche o da fonti amministrative. [26]

Come già analizzato in precedenza, il principio di limitazione delle finalità[27] viene in aiuto alla necessità di garantire la liceità del riutilizzo dei dati. Il GDPR prevede un test di compatibilità tra la finalità del trattamento precedente e quello che si vuole attuare, che trova la sua fonte all'articolo 6. Nel paragrafo 4, è previsto che il titolare del trattamento dei dati deve tenere conto di "*a) ogni nesso tra le finalità per cui i dati personali sono stati raccolti e le finalità dell'ulteriore trattamento previsto, b) del contesto in cui i dati personali sono stati raccolti, c) della natura dei dati personali, d) delle possibili conseguenze dell'ulteriore trattamento previsto per gli interessati, e) dell'esistenza di garanzie adeguate, che possono comprendere la cifratura o la pseudonimizzazione*".

Sebbene la normativa possa sembrare chiara, non sono stati implementati metodi di controllo e verifica dell'effettiva applicazione di questo test di compatibilità. Il ricercatore, quindi, potrebbe trovarsi in una situazione complicata caratterizzata da forte incertezza circa la liceità del trattamento. Questo, purtroppo, risulta un elemento negativo per il ricercatore europeo che potrebbe non riuscire a prendere parte a contesti di ricerca internazionali fortemente competitivi e dominati da tempi brevi e risultati immediati[28].

---

[25] http://www.dst.dk/

[26] Pasquetto, I V, et al.. *"On the Reuse of Scientific Data*". Data Science Journal, vol. 16, no. 0, 2017

[27] Articolo 5, lettera b, Regolamento UE 2016/679.

[28] E.B. Van Veen, *Observational health research in Europe: understanding the GPDR and underlying debate*, in *European Journal of Cancer*, 104, 2018.

Il riutilizzo dei dati per fini di ricerca non richiede solo una cornice normativa forte e coerente, ma anche un'infrastruttura tecnica in grado di integrare dati provenienti da fonti differenti, in particolare quelle amministrative.

I dati amministrativi sono dati che vengono raccolti dalle istituzioni pubbliche nell'esercizio delle loro funzioni amministrative. Essi possono riguardare la popolazione, l'occupazione, l'istruzione, la giustizia. Vengono raccolti attraverso censimenti, dichiarazioni fiscali, sistemi informativi e hanno un grandissimo potenziale nella valutazione dell'attuazione di politiche pubbliche.

Nel contesto europeo, le esperienze nazionali svolgono un ruolo fondamentale per comprendere il livello di implementazione di questi sistemi. In Italia, l'ISTAT ha messo in atto un'infrastruttura centralizzata per la gestione dei dati amministrativi, con l'obiettivo di rafforzarne l'utilizzo e ottimizzarne il processo di produzione. Il SIM, cioè Sistema Integrato di Microdati, assolve proprio a questa funzione, costituendo un deposito per i dati amministrativi acquisiti dall'ISTAT[29].

Nato nel 2013, questo sistema lavora attraverso tre unità di base (Individui, Unità Economiche, Luoghi) che subiscono un processo di integrazione non solo tra elementi dello stesso tipo, ma anche di tipi diversi, fornendo la possibilità di creare un profilo molto dettagliato che ha potere informativo elevato. Ad esempio, si possono combinare le iscrizioni di un individuo agli anni scolastici, oppure il reddito da lavoro con il numero di componenti familiari.

Per poter garantire il rispetto della normativa vigente in materia di trattamento dei dati, l'ISTAT promuove l'elaborazione di alcune procedure designate ad adempiere alle finalità di *privacy by design* e *by default*. Innanzitutto, si opera una distinzione in ingresso tra dati personali e non. I primi subiranno un processo di pseudonimizzazione, che si compone di due fasi: nella prima si opererà una separazione degli identificativi presenti nel dataset dalle variabili e l'assegnazione di codici identificativi validi nel tempo, mentre nella seconda avviene l'abbinamento effettivo tra la variabile e quelle già presenti all'interno del dataset[30].

[29] M.C. Runci, G. Di Bella, L. Galiè, *Il sistema di integrazione dei dati amministrativi in ISTAT* in *ISTAT working papers* n.18, 2016

[30] G. Di Bella, *Il Sistema di Documentazione dei Dati Amministrativi in ISTAT*, 2021, ISTAT.

Una volta assicurata la privacy degli interessati ed aver integrato i nuovi dati con quelli già presenti, si produce un output che viene reso disponibile agli utenti interni dell'ISTAT a fini statistici. Combinare le informazioni statistiche per fornire una visione unitaria dei fenomeni facilita la lettura integrata delle informazioni e garantisce risultati molto più accurati. La realizzazione di questi sistemi ad alto livello di integrazione necessita, tuttavia, di un elevato tasso di centralizzazione, che potrebbe scontrarsi con il ritardo tecnologico che caratterizza la Pubblica Amministrazione in Italia.

2.3. Tecniche di pseudonimizzazione dei dati

Un concetto fortemente enfatizzato dal GDPR e più volte nominato nel corso di questa tesi è l'importanza della "Privacy by Design"[31], un approccio alla privacy che consiste nell'integrare misure di protezione dei dati personali fin dall'inizio di un progetto, piuttosto che aggiungerle successivamente.

Il principale metodo di implementazione di questo concetto è la pseudonimizzazione, definita anch'essa dal GDPR come "*il trattamento dei dati personali in modo tale che i dati personali non possano più essere attribuiti a un interessato specifico senza l'utilizzo di informazioni aggiuntive, a condizione che tali informazioni aggiuntive siano conservate separatamente e soggette a misure tecniche e organizzative intese a garantire che tali dati personali non siano attribuiti a una persona fisica identificata o identificabile*"[32]. Pseudonimizzare significa, quindi, rendere l'informazione meno accessibile a utenti non autorizzati.

I microdati, oggetto del presente studio, contengono spesso identificatori diretti da sostituire o, talvolta, eliminare. Richiamando le categorie di microdati descritte in precedenza, si distinguono diversi livelli di protezione del dato. I file ad uso pubblico sono completamente anonimizzati e, per questo, non sono coinvolti dalla normativa privacy. I *secure-use files*, all'opposto, non sono stati sottoposti ad alcun tipo di processo de-identificativo quindi non possono essere utilizzati a fini di ricerca se non subiscono

[31] Articolo 25, Regolamento UE 2016/679.
[32] Articolo 4, Regolamento UE 2016/679.

un'operazione di pseudonimizzazione che li rende file per la ricerca, in cui il rischio di identificazione dell'unità statistica è fortemente ridotto[33].

Nella protezione dei microdati, è necessario implementare tecniche di pseudonimizzazione utilizzando i metodi efficaci e sicuri propri della computer science, invece di semplificare, manipolare o eliminare i dati, come avviene di solito nel controllo della divulgazione statistica. Tra le principali tecniche, è opportuno menzionare il *Random Number Generator* (RNG), l'*hashing* e il *Message Authentication Code* (MAC).

La tecnica più semplice è quella del generatore di numeri casuali (RNG), un meccanismo che produce valori in set che hanno la stessa probabilità di essere selezionati e, quindi, imprevedibili. Il numero casuale selezionato diventa un identificatore anonimo, lo pseudonimo, e viene sostituito all'identificatore diretto. L'RNG assicura la protezione dei dati ma presenta un rischio di collisione: è possibile, infatti, che a due identificatori sia associato lo stesso pseudonimo.

L'*hashing* è una funzione matematica che riceve in input un set di dati di lunghezza variabile e restituisce un output di lunghezza fissa. La funzione è progettata in modo tale che qualsiasi modifica ai dati di input, anche la più piccola, produca un output completamente diverso. Questa tecnica è resistente alla collisione perché è praticamente impossibile trovare due input diversi che producono lo stesso output, così come è difficile risalire all'input analizzando l'output. Una funzione di hash viene applicata direttamente all'identificatore del dato per ottenere lo pseudonimo. Il dominio dello pseudonimo dipende dal risultato prodotto dalla funzione.

La tecnica di *hashing* viene utilizzata per preservare la privacy nelle tecniche di "linkage" dei dati, con il fine di ottenere database perfettamente integrati e con identificatori univoci e sicuri[34].

Il *Message Authentication Code* (MAC) è una tecnica di crittografia che consiste nel generare un tag unico per un messaggio utilizzando una chiave segreta condivisa tra il mittente e il destinatario. Il mittente crea il MAC combinando il messaggio e la chiave segreta usando un algoritmo di crittografia, che può essere l'hash. Questo crea una serie unica di bit che vengono poi intestati al messaggio. Tramite questa tecnica, è possibile

[33] A. Bujnowska, *Access to European Statistical System Microdata* in N. Crato, P. Paruolo, *Data Driven Policy Impact Evaluation*, SpringerOpen, 2019.

[34] Vedi nota 33.

generare uno pseudonimo univoco per ogni entità coinvolta calcolando il MAC sull'identificatore dell'individuo[35].

È importante sottolineare, inoltre, che molti sistemi di integrazione dei dati, tra i quali il SIM nel contesto italiano, utilizzano tecniche di integrazione dei dati di tipo deterministico, in cui l'identificatore appare in tutti i database con lo stesso pseudonimo. L'obiettivo è quello di ottenere un collegamento diretto e accurato tra le informazioni, privo di ambiguità o incertezza sulle unità collegate.

L'ambito di applicazione di queste tecniche di pseudonimizzazione si trova ancora ad uno stato evolutivo arretrato e non è possibile definire un vero e proprio stato dell'arte su queste tecniche. Per questo, Commissione Europea e le altre istituzioni afferenti ad essa dovrebbero cooperare con le comunità di ricerca e le industrie per definire alcune linee guida tecniche e promuovere la pubblicazione di best practices.

[35] ENISA, *Pseudonymisation techniques and best practices*, Novembre 2019.

Capitolo 3.

Casi di studio sull'utilizzo dei microdati

Sommario: 3.1. L'esperienza francese: un gestore, molti produttori – 3.2. L'esperienza danese: il potenziale dei registri amministrativi – 3.3. Privilegiare le fonti amministrative o pluralità e integrazione: pro e contro nell'analisi statistica

3.1. L'esperienza francese: un gestore, molti produttori.

In Francia, l'istituto nazionale di statistica è l'*Institut national de la statistique et des ètudes èconomiques* (INSEE). Questo è gestito dal Ministero dell'Economia e delle Finanze ed è l'unico ente francese che si occupa di gestire la statistica nazionale.

L'accesso ai dati confidenziali esiste dal 1984, quando la legge statistica nazionale[36] ha permesso di consultare questo tipo di dati legati, però, solo alle aziende. Chiunque intendesse servirsi di tale dati doveva presentare un progetto ad una commissione apposita, la quale elaborava un giudizio riguardo alla richiesta, basato principalmente sull'interesse pubblico della richiesta e sulle garanzie di sicurezza forniti dall'istituzione che gestiva il progetto. Storicamente, la commissione ha sempre permesso l'accesso ai dati a coloro che se ne servissero per scopi di ricerca e che prevedessero di pubblicare i risultati ottenuti. I dati venivano consegnati in un supporto mobile, quale un CD-Rom o un DVD[37].

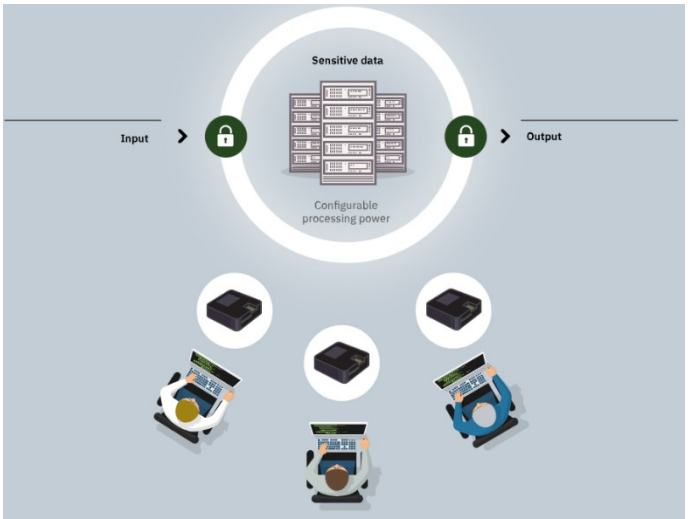

Figura 1. Fonte: CASD official website

Le poche possibilità per il ricercatore di accedere a dati di tipo confidenziale erano, quindi, quelle di ottenere l'accesso agli stessi in maniera temporanea (sebbene questo succedesse raramente) oppure di ottenere output elaborati dall'INSEE, che si componevano di sole tabelle, peraltro molto costose.

[36] Legge n.51-711 del 7 Giugno 1951 (modificata) sul segreto statistico.
[37] Jean-Francois Royer, *Microdata access for researchers in INSEE-France*, Cardiff, 2009.

Nel 2008, la legge statistica viene modificata e viene fornito l'accesso ai dati sugli individui e sulle famiglie, ma solo per scopi di ricerca scientifica o statistici. Chiaramente, questa ammenda ha dato vita ad un nuovo tipo di problema: creare un *data center* per la ricerca che contenga dati confidenziali.

A tale scopo, viene istituito il *Centre d'Accès Sécurisé aux Données* (CASD). Creato come un'infrastruttura dedicata per consentire l'accesso sicuro ai dati sensibili per scopi di ricerca e analisi, garantisce al contempo la protezione della privacy e la prevenzione della divulgazione non autorizzata. Confrontando la propria esperienza con quella di altri paesi[38], la Francia sviluppa un proprio sistema di accesso remoto ai microdati provenienti da statistiche ufficiali.

Il CASD ha progettato un box per computer, denominato SD-Box, attraverso il quale l'utente può effettuare il login ed avere accesso remoto agli strumenti di trattamento dei dati, che in tal modo rimangono in un ambiente protetto, detto bolla sicura. Per l'esportazione degli stessi, è necessario richiedere l'autorizzazione[39].

Avere una tecnologia che garantisce l'autenticazione degli utilizzatori e la tracciabilità dei dati fornisce garanzie essenziali per una diffusione sicura dei dati sensibili. Proprio per questa ragione, il ricercatore non può utilizzare propri strumenti di analisi dei dati, ma deve avvalersi di quelli resi disponibili nel server del CASD, che, fin dalla fase di progettazione, si occupa di mantenere aggiornati e disponibili.

Dopo aver ottenuto l'autorizzazione, i ricercatori devono seguire un incontro di formazione presso il CASD durante il quale vengono informati sulle normative vigenti in tema privacy e riservatezza statistica. Sono presentate anche delle regole di sicurezza informatica molto restrittive: l'accesso è personale e va inserita la carta d'identità all'interno del box, che si deve trovare in una stanza chiusa a chiave in cui lo schermo sia visibile solo dall'utente autorizzato.

Il punto di forza della tecnica francese risiede nella capacità di integrare facilmente diversi registri. Come riportato nel sito ufficiale del CASD[40], oltre 500 fonti di dati vengono rese

---

[38] Ad esempio, il Sistema ad accesso remoto NORC *Data Enclave* sviluppato a Chicago. Vedi Lane&Shipp, *Using a remote access data enclave for data dissemination*, International Journal of Couration, 2007.

[39] Conferenza degli Statistici Europei, *Access to French Statistical Data*, 28-30 Ottobre 2013, Canada.

[40] https://www.casd.eu/en/#casd

disponibili nel server. Garanzie di sicurezza e collaborazione sono alla base della sinergia tra CASD e produttori di dati, permettendo una notevole riduzione dei costi di investimento nello sviluppo di software sicuri e *compliant* con la normativa privacy. Proprio per questa ragione, la pubblicazione del GDPR non ha creato disagio al sistema statistico francese, già addestrato a lavorare con standard qualitativi e di sicurezza elevati. Le garanzie di sicurezza non sono diventate altro che requisiti legali[41].

3.2. L'esperienza danese: il potenziale dei registri amministrativi

In Danimarca, l'agenzia responsabile per la gestione e la distribuzione dei microdati è *Statistics Denmark* (DS). Il compito principale di questa agenzia è la raccolta dei dati da registri amministrativi. Similmente all'ISTAT, poi, questi dati vengono analizzati per produrre statistiche che riflettono lo stato socio-economico del paese.

Nel contesto della distribuzione dei microdati, l'ente ha adottato rigide misure di sicurezza e accesso ai dati per garantire l'integrità delle informazioni personali in essi contenute. Innanzitutto, il richiedente l'accesso ai dati, sia esso un ricercatore singolo o un complesso di ricerca, deve compilare un *form* nel sito web dell'agenzia. Agli ambienti di ricerca e analisi finanziati con fondi pubblici e alle fondazioni di beneficenza, è garantita l'autorizzazione, perché frutto di un accordo tra *Statistics Denmark* e *Danish e-infrastructure Cooperation*, un'organizzazione che si occupa di garantire la presenza di infrastrutture per la ricerca e l'insegnamento. Nel settore privato, possono essere autorizzate organizzazioni non-governative e società di consulenza, a patto che questi non ottengano l'accesso ai dati delle aziende.

Il focus dell'ente statistico in questa fase di approvazione dei ricercatori risiede nell'analisi approfondita delle tecniche di sicurezza implementate e al numero di persone con le capacità di gestire grandi volumi di dati e registri. Nell'ambiente, dev'essere garantita la presenza di un direttore delle risorse umane che supervisioni l'accesso dei dipendenti ai microdati e garantisca che questi siano a conoscenza delle regole di accesso agli stessi. Ad integrazione di questo requisito, ci devono essere almeno tre persone che abbiano maturato una buona esperienza nella gestione di grandi database e abbiano una solida conoscenza delle regole di sicurezza.

---

[41] Commissione Nazionale Informatica e Libertà (CNIL), *Regime giuridico applicabile ai trattamenti a fini di ricerca scientifico (non sanitario)*, 2019.

L'autorizzazione viene approvata dal Direttore Generale, che pone particolare attenzione alle competenze dell'interessato nella gestione dei microdati e alla conoscenza delle norme di sicurezza. Se la richiesta viene approvata, viene stipulato un accordo di accesso ai microdati che delinea i termini e le condizioni, definendo i confini dell'uso e della divulgazione del materiale utilizzato.

I file di microdati per i quali il ricercatore ottiene l'autorizzazione sono resi disponibili in un server di ricerca implementato da *Statistics Denmark*, in maniera completamente anonimizzata. È obbligatorio presentare un progetto di ricerca che va inviato alla divisione dell'ente statistico che si occupa dei servizi di ricerca, la quale individua le modalità migliori di trasmissione dei dati.

Per l'accesso al servizio, il ricercatore deve firmare un accordo che garantisce la protezione e la segretezza dei dati. Inoltre, stabilisce che tutto il lavoro di ricerca deve svolgersi all'interno del server e che non devono essere effettuati tentativi di identificazione di persone o imprese. L'esportazione di microdati è considerata una severa violazione dell'accordo, in quanto possono essere rimossi dal server solo dati aggregati, dei quali non è possibile l'identificazione. I risultati elaborati possono essere salvati in un file specifico del server, che verrà inoltrato via e-mail al ricercatore. In questo modo, *Statistics Denmark* riesce a garantire il controllo dei dati esportati.

La Danimarca può essere considerata come il leader nella raccolta di informazioni personali che riguardano i cittadini in registri. I dati contenuti nel *Det Centrale Personregister*, cioè il registro centrale delle informazioni sulle persone, venne implementato 50 anni fa e contiene un identificatore unico per ogni cittadino danese. Attraverso questo numero, *Statistics Denmark* può accedere all'identificazione diretta ed integrare diversi registri, anche se afferenti ad aree sociali diverse, come il mercato del lavoro e l'educazione[42].

Il governo danese, ad esempio, richiede ad ogni cittadino ed impresa registrata di dotarsi di un conto bancario personale, che verrà etichettato come *NemKonto*[43]. Tramite questo strumento, tutte le transazioni tra cittadino, impresa e pubblica amministrazione sono registrate elettronicamente e possono essere utilizzate come input per la produzione di

[42] http://www.nordiclabourjournal.org/i-fokus/in-focus-2018/theme-nordic-statistics/article.2018-12-17.2801676342

[43] https://www.nemkonto.dk/servicemenu/engelsk

nuova statistica. In questo modo, è possibile ottenere informazione statistica sul mercato del lavoro, come un indicatore che permette di individuare tutti i percettori di stipendio.
Niels Ploug, direttore delle statistiche sociali presso *Statistics Denmark*, riconosce il primato della Danimarca nella registrazione dei dati, ma individua anche delle debolezze per quanto riguarda i dati quantitativi. L'ente danese non si occupa di effettuare molti sondaggi o indagini, per questo non possiede dati riguardo a considerazioni soggettive sulle condizioni lavorative dei lavoratori danesi, come lo stress o la frequenza del congedo per malattia.

3.3. Privilegiare le fonti amministrative o pluralità e integrazione: pro e contro nell'analisi statistica

Sono stati analizzati due approcci alla stessa materia che hanno seguito percorsi molto diversi tra loro. Nel primo caso, l'INSEE ha sviluppato una tecnica di accesso ai microdati da remoto che avviene presso centri ad hoc, ai quali confluiscono i dati amministrativi e quelli di altri enti. Nel secondo caso, DS ha preferito imporsi come produttore e gestore della statistica ufficiale nazionale sviluppando tecniche di accesso che si fondano su basi di dati provenienti da registri amministrativi.
Lavorare con un sistema che favorisce l'utilizzo di una sola fonte può garantire attendibilità e completezza dei dati. I dati amministrativi raccolti da enti e istituzioni possono essere considerati affidabili e completi, poiché raccolti per scopi amministrativi e di monitoraggio. Considerate le finalità di ricerca per le quali vengono trattati, assicurare al ricercatore che la conoscenza che può trarre dalla manipolazione dei dati possa risultare veritiera è una forte garanzia di attendibilità dei risultati. D'altra parte, è fondamentale prendere in considerazione le finalità della ricerca. L'uso esclusivo di fonti amministrative rischia di non coprire la totalità degli aspetti e delle variabili di interesse, limitando la gamma di dati disponibili e minando la qualità dei risultati della ricerca.
L'utilizzo delle sole fonti amministrative assicura vantaggi anche in termini economici: avvalendosi di dati già "pronti", si evita la necessità di condurre indagini specifiche, riducendo così il carico di lavoro e i costi associati alla raccolta dei dati. Tuttavia, i dati amministrativi potrebbero non fornire informazioni dettagliate su determinati fenomeni sociali, rendendo difficile l'analisi approfondita, che richiede l'integrazione con sondaggi.

L'implementazione del modello francese, invece, trova i suoi punti di forza nell'integrazione delle fonti generando vantaggi sia per i produttori di dati che per i ricercatori.

Dal punto di vista dei produttori dei dati, è necessario ricordare che la maggior parte di questi non svolge come compito principale quello della produzione dei dati, ma li rende ugualmente disponibili ai fini della ricerca scientifica. Il fatto che vi sia un terzo, in questo caso il CASD, che garantisce la diffusione sicura dei dati permette al produttore di risparmiare ingenti somme di denaro che andrebbero investite per creare infrastrutture per fornire questo servizio. Anche il CASD svolge un ottimo lavoro di ottimizzazione perché può avvalersi dello stesso tipo di servizio per tutti i produttori dei dati, generando un buon risparmio nei costi operativi. Inoltre, non è più necessario che il ricercatore stipuli un accordo specifico con ogni produttore dei dati, ma un unico contratto tra ricercatore e CASD regola l'intero rapporto.

Questo sistema offre anche il grande vantaggio per i ricercatori di integrare diverse fonti per l'uso congiunto o l'abbinamento nello stesso ambiente di lavoro. Avvalendosi di un'ampia gamma di dati disponibili, il ricercatore può permettersi di svolgere analisi molto approfondite ed interdisciplinari, coprendo un maggior numero di fenomeni. Gestire e integrare una vasta gamma di fonti di dati, tuttavia, può essere complesso e richiede competenze significative per garantire integrità ai dati, in particolare delle tecniche di pulizia. Inoltre, anche se il CASD mette a disposizione oltre 500 fonti di dati, potrebbero ancora mancare alcune fonti che contengono i dati di interesse. Abbonarsi a CASD costa ad oggi 253,00€ al mese[44] per il pacchetto base che include l'erogazione dei servizi per 1 utente: questo prezzo potrebbe risultare proibitivo per un ricercatore singolo che si abbona ma al quale non è garantita la presenza di tutti i dati che necessita.

Come già accennato, la finalità della ricerca può essere identificato come il fattore discriminante tra la maggiore efficienza di un modello piuttosto che dell'altro. Ad esempio, si immagini un'indagine che valuta l'impatto delle politiche di istruzione sulla partecipazione degli studenti alle scuole superiori. In questo caso, potrebbe essere sufficiente utilizzare registri amministrativi forniti dal Ministero dell'Istruzione, che contengono informazioni su studenti, scuole e politiche di istruzione. In questo modo è

[44] https://www.casd.eu/en/tarifs-2/

possibile analizzare il numero di studenti iscritti in diverse scuole e misurare fattori importanti quali il tasso di abbandono scolastico e i risultati degli esami.

Nel caso, invece, in cui si voglia studiare la correlazione tra obesità infantile e fattori socio-economici della famiglia, potrebbe essere necessario integrare diverse fonti di dati per comprendere il fenomeno al meglio. Ad esempio, i fattori legati alla salute dell'infante vengono estratti da dati amministrativi dei servizi sanitari. I dati alimentari degli alimenti e i dati ambientali possono essere forniti di agenzie che si occupano di indagini rispettivamente nutrizionali e ambientali. L'integrazione di queste informazioni permette di comprendere al meglio la capacità di accesso a cibo sano, le condizioni economiche delle famiglie e gli ambienti di vita.

Conclusioni

Sommario: 1. Confronto situazione italiana-europea – 2. Possibili sviluppi futuri

1. Confronto situazione italiana – europea

Da un punto di vista tecnico, le modalità di accesso ai microdati implementate finora nel contesto italiano sono molto limitate e obsolete. L'ISTAT mette a disposizione il Laboratorio per l'Analisi dei Dati ELEmentari (ADELE) per consultare i file di microdati relativi a tutte le rilevazioni statistiche dell'Istat. Questi dati sono, in via preventiva, privati degli identificativi diretti.

Per accedere al laboratorio ADELE, è necessario essere riconosciuti come "ente di ricerca" dal Comstat o da Eurostat. All'interno di questo laboratorio è possibile predisporre analisi statistiche sui dati elementari raccolti direttamente dall'Istat che riguardano individui, famiglie, imprese e istituzioni. Vengono messi a disposizione alcuni strumenti software per l'analisi. Per quanto riguarda l'esportazione dei risultati, è necessario ottenere una valutazione di sicurezza positiva dell'output dagli esperti del Laboratorio[45].

L'accesso a questi dati può avvenire solo in laboratori dotati di postazioni fisiche predisposte direttamente dall'ISTAT presso le sedi regionali e provinciali. Pertanto, non è ancora stata implementata alcuna forma di consultazione dei file da remoto, a differenza della maggioranza degli altri paesi europei.

All'interno del laboratorio, inoltre, è possibile lavorare su dati elementari cui sono stati applicati metodi di controllo per la tutela della riservatezza da parte dell'ISTAT. A differenza delle altre esperienze europee sopra descritte, non è presente alcun tentativo di utilizzare registri provenienti da fonti amministrative, e tantomeno di integrarli con quelli già presenti. Se un ricercatore necessitasse, ad esempio, di accedere a registri di dati sul lavoro, dovrebbe rivolgersi direttamente al Ministero del Lavoro per richiederne l'accesso, rendendo la procedura molto costosa sia in termini economici che temporali, nonché inutilmente complicata[46].

Realizzare un sistema di archivi amministrativi sembra ancora molto lontano dall'attuale percorso italiano. Ma anche se sorgesse una propensione verso l'integrazione di dati

[45] ISTAT, Linee guida per l'accesso ai microdati dell'Istat, Roma, Maggio 2022.

[46] http://dati.lavoro.gov.it/microdati-la-ricerca

amministrativi, quale sarebbe la tecnica migliore? Gli archivi sono numerosi e molto diversi tra di loro. È bene garantire la presenza di un'istituzione unica che regola la produzione e la distribuzione dei microdati, come avviene in Danimarca, o seguire l'esempio francese tentando di armonizzare diversi produttori di dati definendo degli standard che permettano di costruire un solo portale d'accesso?[47]

2. Possibili sviluppi futuri

Il percorso verso un utilizzo integrato e condiviso, almeno a livello europeo, dei microdati sembra un obiettivo di difficile raggiungimento. Tuttavia, dopo la promulgazione del GDPR, la ricezione e la volontà di adattamento alle norme della maggior parte dei paesi membri sembra far intendere che, seppur possa essere una sfida impegnativa, l'impegno e la determinazione costituiranno un aiuto essenziale per il compito svolto dai ricercatori. Dopo l'entrata in vigore del GDPR, sembra che sia stata posta particolare attenzione alla definizione delle modalità del trattamento dei dati personali ex-ante, concentrandosi sull'adozione di misure preventive e sulla valutazione anticipata delle implicazioni per la privacy. È importante, però, riconoscere che la gestione ex-post è altrettanto fondamentale per poter affrontare al meglio eventuali problematiche che potrebbero sorgere durante il ciclo di vita dei dati. Questo aspetto richiede, tra le altre, buone capacità di implementazione di meccanismi di monitoraggio che permettano la consultazione dello storico degli accessi ai dati.

A tale scopo, potrebbe essere esplorata l'applicazione della tecnologia blockchain per costruire un registro di accesso. Questo tipo di tecnologia costituisce un registro immutabile e distribuito che registra le transazioni, nel nostro caso gli accessi, in blocchi. Questi creano una catena, alla quale vengono aggiunte nuove informazioni in ordine cronologico. Ogni blocco contiene un hash crittografico unico che verifica l'integrità dei dati al suo interno[48]. In questo modo, è possibile risalire ad ogni accesso ai dati, identificando l'utente, la data e l'ora. Questa tracciabilità è utile per scopi legali, in particolare in quei casi in cui avvengano manipolazioni delle informazioni o vengano violati i vincoli di segretezza a cui sono sottoposti i ricercatori per accedere ai dati. Questo

[47] U. Trivellato, *Accesso ai micordati, ricerca scientifica e valutazione delle politiche: urge un cambio di passo*, in EticaEconomia, Ottobre 2019.

[48] Brookshear, J. Glenn, Brylow, Dennis, *Informatica: una panoramica generale*, Milano, Pearson, 2016.

tipo di rischio è molto elevato perché una volta che i dati sono "allo scoperto" possono essere copiati senza limiti, a costo marginale quasi nullo. Diventa, quindi, impossibile rintracciarli.

L'impiego della tecnologia blockchain per tenere un registro di accesso ai dati sicuro e immutabile. All'interno di ogni rete, solo utenti autorizzati possono aggiungere i blocchi e visualizzare le informazioni contenute in essa. Questo garantisce che nessuna entità abbia il controllo dell'intera rete, rendendola resistente alle frodi.

Tuttavia, l'implementazione di registri distribuiti presenta molti problemi legati alla scalabilità. La necessità di elaborare un gran numero di informazioni può influire sulla velocità delle operazioni perché ogni nodo deve verificare autonomamente ogni transazione. Il rischio è, quindi, che, se il sistema viene implementato a livello nazionale o europeo, si verifichi una situazione simile a quella che ha dovuto affrontare Bitcoin nel 2017, il quale ha rischiato il collasso per il numero di transazioni troppo elevato da gestire.

L'applicazione della tecnologia blockchain nel contesto dei microdati e della protezione della privacy rappresenta una prospettiva interessante per il futuro della ricerca scientifica. È importante che vengano condotti studi e collaborazioni interdisciplinari per esplorare a fondo questa possibilità, così da definire linee guida e *best practices* che promuovano un utilizzo responsabile e sicuro dei microdati nel rispetto dei principi della privacy.

Bibliografia

A. Bujnowska, *Access to European Statistical System Microdata* in N.Crato, P.Paruolo, *Data Driven Policy Impact Evaluation*, SpringerOpen, 2019.

American Law Institute, "*4 categories: intrusion into seclusion, appropriation of name or likeness, public disclosure of private facts* and *placing a person in a false light"*, *Restatement (2nd) of Torts*, §652

Brookshear, J. Glenn, Brylow, Dennis, *Informatica: una panoramica generale*, Milano, Pearson, 2016.

Carta dei Diritti Fondamentali dell'Unione Europea, *Articolo 13 "Libertà delle arti e delle scienze"*, 2000.

Commissione europea, *Che cosa disciplina il regolamento generale sulla protezione dei dati?"*, https://commission.europa.eu/law/law-topic/data-protection/reform/what-does-general-data-protection-regulation-gdpr-govern

Commissione europea, *Che cosa disciplina il regolamento generale sulla protezione dei dati?"*, https://commission.europa.eu/law/law-topic/data-protection/reform/what-does-general-data-protection-regulation-gdpr-govern

Commissione nazionale informatica e libertà (CNIL), *Regime giuridico applicabile ai trattamenti a fini di ricerca scientifico (non sanitario)*, 2019.

Comunicato Stampa della Commissione Europea, *Dati nell'UE: forte impegno della Commissione per aumentare la disponibilità dei dati e promuoverne la condivisione nel campo dell'assistenza sanitaria*, Bruxelles, 2018.

Conferenza degli Statistici Europei, *Access to French Statistical Data*, 28-30 Ottobre 2013, Canada.

Convenzione Europea dei Diritti dell'Uomo, Articolo 8 "Diritto al rispetto della vita privata e familiare ", 1950

Cuffaro, Vincenzo, et al. *I dati personali nel diritto europeo*. G. Giappichelli, 2019, 573-578.

D. Muller & J. Moller, *Giving the International Scientific Community Access to German Labor Market Data: A Success Story* in *Data-Driven Policy Impact Evaluation: How Access to Microdata is Transforming Policy Design*, Germania, Springer International Publishing, 2018.

D. Solove, *Conceptualizing Privacy*, 90 California Law Review 1087, 2002.

ENISA, *Pseudonymisation techniques and best practices*, Novembre 2019.

E.B. Van Veen, *Observational health research in Europe: understanding the GPDR and underlying debate*, in *European Journal of Cancer*, 104, 2018.

G. Di Bella, *Il Sistema di Documentazione dei Dati Amministrativi in ISTAT*, 2021, ISTAT.

GDPR, *Considerando n.157*, 2016.

GDPR, *Considerando n.4*, 2016.

GDPR, *Articolo 4 "Definizioni"*, 2016.

GDPR, *Articolo 5 "Principi applicabili al trattamento di dati personali"*, 2016.

GDPR, *Articolo 25 "Protezione dei dati fin dalla progettazione e protezione dei dati per impostazione predefinita"*, 2016.

ISTAT, *Linee guida per l'accesso ai microdati dell'Istat*, Roma, Maggio 2022.

ISTAT, *File di Microdati*, 2023. https://www.istat.it/it/dati-analisi-e-prodotti/microdati

Jean-Francois Royer, *Microdata access for researchers in INSEE-France*, Cardiff, 2009.

Lane&Shipp, *Using a remote access data enclave for data dissemination*, International Journal of Couration, 2007.

Legge n.51-711 del 7 Giugno 1951 (modificata) sul segreto statistico.

M.C. Runci, G. Di Bella, L. Galiè, *Il sistema di integrazione dei dati amministrativi in ISTAT* in *ISTAT working papers* n.18, 2016

M. Preisler, *Denmark leads the way on statistics using microdata* in *Nordic Labour Jornal*, 17 Dicembre 2018. http://www.nordiclabourjournal.org/i-fokus/in-focus-2018/theme-nordic-statistics/article.2018-12-17.2801676342

Open Data, *Microdati per la ricerca*, Ministero del Lavoro e delle Politiche Sociali, 2023. http://dati.lavoro.gov.it/microdati-la-ricerca

P. Bisogno, *Enciclopedia Italiana Treccani*, IV Appendice, 1981.

P. Pasquetto, I V, et al.. *"On the Reuse of Scientific Data*". Data Science Journal, vol. 16, no. 0, 2017

Regolamento (CE) n.223 del Parlamento Europeo e del Consiglio, *Articolo 19*, 2009.

Regolamento (UE) n.557 della Commissione Europea, *Articolo 2*, 2013.

S.Ruggieri, *Towards a Digital Ecosystem of Trust: Ethical, Legal and Societal Implications*, in *Opinio Juris In Comparatione*, 1, 2021.

S. Stefanelli, *Ricerca scientifica e privacy. I limiti europei sono sufficienti. Perché l'Italia vuole andare oltre?*, in *Quotidiano Sanità*, 2018.

Sito online di *Statistics Denmark* http://www.dst.dk/

Sito online dell'INS francese. https://www.casd.eu/en/#casd

Sito online del servizio svedese Nemkonto https://www.nemkonto.dk/servicemenu/engelsk

U. Trivellato, *Data-Driven Policy Impact Evaluation: How Access to Microdata is Transforming Policy Design*, Germania, Springer International Publishing, 2018.

U. Trivellato, *Accesso ai micordati, ricerca scientifica e valutazione delle politiche: urge un cambio di passo*, in EticaEconomia, Ottobre 2019.

Warren and Brandeis, *The Right to Privacy*, 1890.