Politecnico di Torino

Corso di Laurea Magistrale in Ingegneria Gestionale
indirizzo Finance

A.a. 2024/2025

Politecnico
di Torino

L'economia dei dati: il ruolo dei data broker

Relatore:
Prof.ssa Laura ABRARDI

Candidato:
Diego BELMONTE

ABSTRACT

In un mondo sempre più digitale e in costante evoluzione, i Big Data sono un asset strategico in quella che ormai viene definita “Data Economy”. Molte aziende, organizzazioni ed enti pubblici e privati guardano ai dati come una fonte inesauribile di sapere e di vantaggio competitivo. L’incessante crescita della digitalizzazione, con il conseguente aumento delle interazioni online, ha fatto sì che il volume di dati crescesse ad un ritmo senza precedenti e ha inoltre amplificato la facilità di raccolta dei dati. Basta, infatti, la registrazione e il login ad un sito web per rilasciare un grande ammontare di informazioni personali. La facilità di raccolta, la velocità di circolazione, il valore intrinseco e i costi di archiviazione contenuti, grazie all’emergere delle nuove infrastrutture di storage come i cloud, rendono il mercato dei dati un settore potenzialmente in grado di garantire entrate regolari e durature nel tempo. In questo contesto, i data broker, ovvero compagnie ed entità commerciali che raccolgono una notevole quantità di dati sui consumatori, si configurano come attori principali. Tali dati, che possono anche essere sensibili, vengono raccolti a partire da varie fonti, sia online che offline. Questa tesi ha lo scopo di porre un focus sul business dei data broker, proponendo una comprensione più profonda delle loro pratiche (tanto poco trasparenti quanto invasive) e delle loro implicazioni sulla privacy dei cittadini.

Nella prima sezione l’elaborato offre una panoramica generale sul mondo dei dati, ne definisce le caratteristiche fondamentali e ne descrive il processo di estrazione di valore che va dalla raccolta all’ immagazzinamento fino all’utilizzo effettivo del dato da parte delle imprese per scopi di discriminazione di prezzo e targeted advertising.

Nella seconda sezione l’elaborato descrive i principali data broker operanti nel mercato, le fonti utilizzate, le tipologie di prodotti offerti e i principali clienti divisi per tipologia di prodotti. Inoltre, viene discusso il concetto di capitalismo di sorveglianza. Infine, viene presentata una breve review sulla letteratura dei modelli economici dei data broker.

Nella terza sezione si discutono le implicazioni della profilazione su larga scala e i conseguenti rischi per la privacy dei cittadini e di discriminazione su base etnica o religiosa. Inoltre, si parlerà del rischio di inaccuratezza e imprecisione dei dati raccolti, che può portare a profilazioni errate e al mancato accesso ad alcuni servizi, come

assistenza sanitaria e credito bancario. Si discuteranno anche le implicazioni del business dei data broker sulla sicurezza nazionale.

Nel quarto capitolo si discuterà dell'importanza della privacy e della protezione dei dati personali, portando come esempio lo scandalo di Cambridge Analytica. In seguito, si discuterà della principale normativa europea vigente che concerne i data broker: il GDPR, insieme ad altre normative più recenti, come Digital Markets Act, Digital Services Act e Data Governance Act, analizzando l'impatto di queste normative sul business dei data broker. Verrà, inoltre, posto l'accento sul diverso approccio alla privacy tra Europa e Stati Uniti. Infine, si discuteranno le prospettive e le sfide future per una regolamentazione più efficace.

INDICE

INTRODUZIONE

I Data Broker sono attori principali nella "Data Economy", un'economia nella quale i dati digitali diventano una risorsa fondamentale, per aziende e organizzazioni, per sostenere l'innovazione e la creazione di valore. I data broker si occupano di raccogliere dati, anche personali, sui consumatori, per poi vendere tali dati a terzi, che possono essere aziende private o pubbliche, entità governative, istituzioni finanziarie, compagnie assicurative e altri data broker. I data broker possiedono grandi volumi di informazioni dettagliate sui consumatori come: dati anagrafici, informazioni demografiche, informazioni sul reddito, informazioni sulla situazione finanziaria, stato di salute, interessi, abitudini d'acquisto, comportamento online ecc. e spesso utilizzano tali informazioni per creare delle "segmentazioni" e dei profili sui consumatori che vengono poi venduti a terze parti. Essi hanno avviato un processo di "commodification" dei dati. In questo contesto i dati personali vengono visti soltanto come un'opportunità di profitto e trattati alla stregua di una merce, avente un valore economico e scambiata (spesso senza la consapevolezza e il consenso degli utenti) per ottenere un profitto. L'operato di questi soggetti solleva importanti preoccupazioni riguardo la privacy e i diritti dei cittadini nonché riguardo la trasparenza con la quale i dati vengono raccolti, trattati e in seguito scambiati sul mercato. Questo elaborato, partendo da una panoramica generale sul mondo dei dati, si propone di approfondire le pratiche, altrettanto pervasive quanto poco trasparenti, dei data broker e il loro ruolo all'interno della "data economy", nonché le implicazioni e i rischi di queste pratiche per la privacy, i diritti e la sicurezza dei cittadini.

Nella prima sezione dell'elaborato viene presentata una panoramica sul mercato dei dati e vengono descritte le caratteristiche essenziali dei Big Data, le cosiddette "cinque V", ovvero: Volume, Velocità, Varietà, Veridicità e Valore. Viene descritto anche il processo di estrazione del valore dai dati, ovvero la "Data Value Chain", che è composta da vari step tra cui: acquisizione, analisi, immagazzinamento e infine utilizzo del dato. Quest'ultimo step rappresenta l'effettiva creazione di valore aggiunto. Alla fine della prima sezione verranno presentati alcuni dei possibili utilizzi dei dati che comprendono: supporto alle strategie di "discriminazione di prezzo" operate dalle imprese, "targeted

advertising" (l'uso dei dati per indirizzare ai consumatori "annunci mirati", ovvero rivolti solamente a specifici gruppi di consumatori individuati sulla base dei dati raccolti su di loro relativamente alle abitudini d'acquisto e comportamento online) e infine uso dei dati come strumento di supporto ai processi decisionali.

Nel secondo capitolo vengono descritti i principali player del mercato dei data broker, come Oracle e Acxiom, le fonti utilizzate per l'approvvigionamento, i prodotti offerti e i principali clienti. Acxiom, ad esempio, pubblicizza oltre 2,5 miliardi di dati sui consumatori di tutto il mondo e gestisce 15 mila database sui consumatori. Le fonti utilizzate dai data broker per raccogliere informazioni sui consumatori sono varie e includono: "fonti pubbliche" (come registri del tribunale, report di proprietà, registri dei veicoli a motore, profili dei social media), "fonti private" (come istituzioni finanziarie, rivenditori al dettaglio, proprietari di siti web e app, altri data broker) e infine "tracciamento online" attraverso i cookies o il browser fingerprinting. I prodotti offerti includono: "prodotti per la mitigazione del rischio" che possono essere utilizzati dai clienti dei data broker per verificare l'identità di una persona o per prevenire le frodi, oppure possono essere utilizzati da banche o assicurazioni per gestire rispettivamente il rischio di credito e il rischio assicurativo, "prodotti per il marketing e advertising" che possono aiutare le imprese a ottenere informazioni più dettagliate sui consumatori e insight utili a migliorare il marketing e il targeting pubblicitario. Questi prodotti trovano mercato in diversi clienti, tra i quali: istituzioni finanziarie, entità governative, imprese operanti nel marketing e advertising, imprese retail, imprese farmaceutiche, aziende nel settore tecnologico, forze dell'ordine, assicurazioni, singoli individui e altri data broker.

La terza sezione si concentra invece sulle principali implicazioni economiche delle attività dei data broker. Tali attività hanno dato vita a un ecosistema, definito "Capitalismo della Sorveglianza" da Zuboff (2019), in cui gli individui sono costantemente "censiti, valutati, classificati e categorizzati" (Christl, W., 2017) e in cui i loro dati personali vengono accumulati e monetizzati attraverso la vendita. I sofisticati algoritmi di data mining utilizzati dai data broker per raccogliere i dati, non solo potenziano le loro capacità di acquisizione dei dati, ma rendono estremamente complicato per gli individui sfuggire alla loro "sorveglianza", soprattutto per quanto riguarda le interazioni online (Reviglio, U., 2022). Le pratiche dei data broker

sottopongono i consumatori a rischi significativi per la privacy e i diritti dei cittadini, dal momento che i dati raccolti spesso sono dati "sensibili" (come dati sulla salute). Ad esempio, i sistemi dei data broker possono subire un data breach e i dati sensibili degli individui sono a rischio di accesso e divulgazione non autorizzata. Inoltre, i dati raccolti possono contenere errori (il consumatore è associato per errore ad un reato penale) e ciò può portare a profilazioni errate e conseguentemente ad un mancato accesso al credito, ad una mancata assunzione o ad un rifiuto per un alloggio pubblico. I "prodotti di ricerca" forniti dai data broker sono fonte di rischi per i diritti civili e la sicurezza dei consumatori, sia perché possono essere utilizzati anche per scopi di stalking o persecuzione, sia perché l'uso eccessivo da parte delle forze dell'ordine può portare ad un arresto errato o limitazioni ingiustificate della libertà personale. Il capitolo esamina infine le implicazioni sulla sicurezza nazionale, poiché i data broker possiedono anche dati su militari, funzionari governativi ecc., spesso con insufficiente "anonimizzazione", che possono finire nelle mani di attori ostili ed essere utilizzati per pianificare azioni che compromettono la sicurezza nazionale (Sherman, 2021).

Infine, nella quarta sezione verranno analizzate le normative europee attualmente in vigore per la tutela della privacy e il loro impatto sul business dei data broker. Si partirà analizzando il caso di Cambridge Analytica, esempio emblematico di violazione della privacy e di uso illegittimo dei dati, che ha messo alla luce l'importanza della privacy e di una regolamentazione sull'uso dei dati. Tale caso ha funzionato da catalizzatore per l'adozione del GDPR (2016) in Europa. Il "Regolamento Generale sulla Protezione dei Dati" (GDPR) ha introdotto nuovi diritti per i cittadini col fine di aiutarli ad avere maggiore controllo sui propri dati e proteggere la loro privacy. Tra questi diritti figurano: il diritto all'oblio, il diritto alla portabilità dei dati, la richiesta di consenso, il diritto di limitazione ecc. Uno dei punti chiave della norma riguarda le linee guida che devono essere seguite dai soggetti che raccolgono i dati ("titolari del trattamento"). Tra questi principi ricordiamo: il concetto di "minimizzazione dei dati", per cui i dati devono essere "pertinenti e limitati a quanto necessario" per lo scopo del trattamento, oppure il principio della "limitazione della conservazione", per cui i dati devono essere conservati solo per il tempo necessario al conseguimento dello scopo per cui sono stati raccolti, dopodiché bisogna procedere alla cancellazione. Si discuterà anche delle normative

europee che compongono la “Strategia Europea sui Dati”, ovvero Data Act, Digital Services Act, Digital Markets Act e Data Governance Act. Il Data Act è importante poiché estende alcuni diritti già presenti nel GDPR (come il diritto alla portabilità) a qualsiasi dato generato da dispositivi e macchine IoT; quindi, tale normativa in molti casi integra e completa il GDPR. Il GDPR è una delle normative più avanzate in termini di protezione della privacy, tuttavia presenta un deficit di “enforcement”, in quanto è difficile applicare le proprie disposizioni alla specificità e complessità del business dei data broker. Le altre normative europee sono sulla giusta direzione, tuttavia non intaccano direttamente il business dei data broker ed è difficile che abbiano un impatto significativo (Ruschemeier, 2023). Inoltre, in questa sezione, viene presentata la differenza tra l’approccio normativo europeo e quello statunitense, che è un approccio meno restrittivo ed è caratterizzato dall’assenza di una regolazione a livello federale, ma piuttosto di “un mosaico di leggi specifiche per il settore” che regolano il trattamento dei dati soltanto in “determinati settori e situazioni” o per alcuni soggetti (Rieke, A. et al., 2016).

Capitolo 1: Il mercato dei dati

Ci troviamo attualmente in un'epoca in cui stiamo assistendo ad una vera e propria rivoluzione digitale, in cui il volume di dati generati e scambiati su Internet, cresce ad un ritmo senza precedenti. La produzione di dati a livello globale ha avuto un importante incremento nel 2020 guidato principalmente dalla pandemia di COVID-19 che, costringendo milioni di persone a casa (con un conseguente aumento dei click sugli annunci, reazioni sui social media, contenuti in streaming ecc.), ha accelerato la digitalizzazione. Si stima che nel 2020 siano stati creati, acquisiti, copiati e consumati oltre 64,2 ZB[1] di dati in tutto il mondo rispetto ai 41 ZB del 2019, registrando un incremento del 56%.

Nel 2024 si stima che la quantità totale di dati sia stata di circa 147 ZB ma la tendenza sembra destinata ad aumentare; infatti, secondo le stime nel 2025 si arriveranno a superare i 160 ZB fino a toccare il tetto dei 181 ZB (International Data Corporation, 2021).

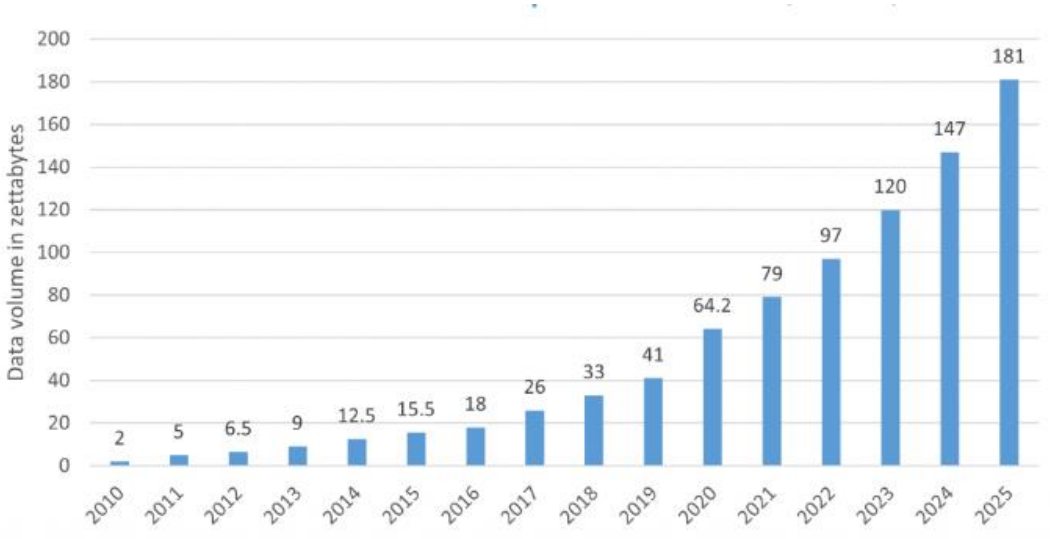

Figura 1: Misura annuale della datasfera globale fonte: (International Data Corporation, Global DataSphere Forecast 2021-2025, 2021)

[1] Zettabytes: 1 ZB= 10^{12} Gigabytes

Oggi, molte aziende, organizzazioni e governi guardano ai Big Data come una fonte di vantaggio competitivo: ad esempio, le aziende possono utilizzare i dati per conoscere meglio le preferenze e i comportamenti dei consumatori basando le loro decisioni strategiche su informazioni concrete. Allo stesso modo, i governi possono utilizzare i Big Data sia per migliorare la qualità dei servizi pubblici e il benessere dei cittadini che per affrontare sfide complesse come le disuguaglianze sociali o il cambiamento climatico. Inoltre, l'analisi dei Big Data offre un vantaggio competitivo alle aziende e migliora l'efficienza dei servizi pubblici aprendo così la strada ad un nuovo tipo di economia detta "Data Economy" (un'economia nella quale le informazioni digitali guidano l'innovazione e la creazione di valore). I Big Data, infatti, stanno trasformando molti settori: dall'assistenza sanitaria ai media, dall'erogazione di servizi energetici ai servizi finanziari e assicurativi. Le aree di applicazione sono moltissime (Cavanillas et al., 2016):

- in ambito sanitario è possibile effettuare una diagnosi più accurata e tempestiva grazie all'analisi di grandi volumi di dati medici e genetici che, dati in pasto ad algoritmi predittivi, identificano le cause di rischio nei pazienti e le tendenze emergenti.
- le banche usano i Big Data per valutare meglio il rischio di credito e prevenire le frodi.
- le compagnie assicurative possono creare prodotti su misura per i clienti migliorando la valutazione del rischio e ottimizzando i premi assicurativi.
- in ambito finanziario e non solo, i Big Data sono alla base di un nuovo paradigma tecnologico, ovvero quello costituito dalla cosiddetta **Blockchain**[2], una sorta di registro distribuito costituito da blocchi (contenti le transazioni) e collegato in più nodi ognuno dei quali mantiene una copia del registro.

[2] La Blockchain è una rete P2P "Peer-to-Peer" dove i nodi comunicano, verificano e approvano le transazioni senza passare da un intermediario centrale (la decentralizzazione delle transazioni riduce i costi e aumenta l'efficienza).

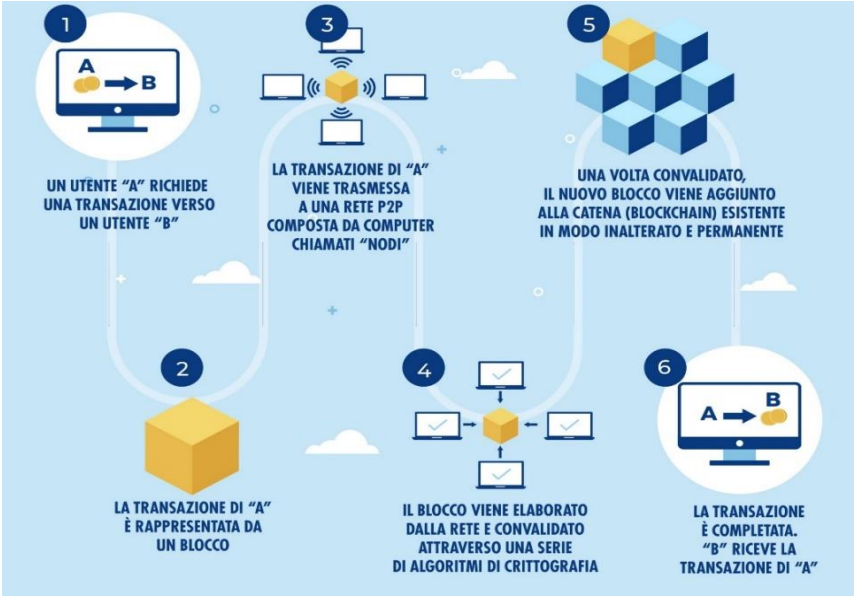

Figura 2: Funzionamento della Blockchain
Fonte: (https://www.udiconer.it/infografica-blockchain-che-cose-e-come-funziona/)

- nel settore retail alcune imprese (ad esempio Amazon) utilizzano i Big Data per analizzare le preferenze e le abitudini dei consumatori col fine di customizzare le offerte, migliorare l'esperienza di acquisto o per migliorare la gestione dell'inventario.
- le piattaforme di streaming, analizzando i dati, possono sfruttare le preferenze dei consumatori per creare contenuti che rispondano meglio ai gusti e alle esigenze degli spettatori migliorando così la fidelizzazione e l'esperienza dell'utente.
- nel mondo del trasporto e della logistica si possono utilizzare i dati per ottimizzare i percorsi, ridurre i tempi di consegna e le emissioni di CO_2.
- nel settore dell'energia l'analisi dei dati aiuta a migliorare l'efficienza e dunque, una corretta programmazione della domanda e dell'offerta.

Nei mercati dei Big Data è possibile trovare delle piattaforme che funzionano secondo la teoria economica dei **mercati a due versanti (two-sided market)**. In genere, tali mercati sono caratterizzati dalla presenza di due gruppi distinti di agenti economici che interagiscono attraverso una **piattaforma online**. Tale piattaforma permette ai due

agenti economici di comunicare tra loro e crea valore aggiunto per entrambi. Essa si finanzia attraverso le fee addebitate ai ristoranti iscritti e alle spese di consegna. Un classico esempio è rappresentato dalle piattaforme di food delivery in cui da un lato, i ristoranti riescono a raggiungere più clienti senza doversi occupare in prima persona della consegna e godono di una maggiore visibilità; dall'altro, i clienti possono attingere da un'ampia selezione di ristoranti e ordinare comodamente da casa.

Un altro aspetto da sottolineare è la presenza di esternalità di rete dirette e incrociate. Quest' ultime implicano che, le decisioni prese dagli agenti appartenenti ad un versante, producono effetti sugli agenti che fanno parte dell'altro versante. Tuttavia, la caratteristica che rende unici questi mercati è che l'asimmetria informativa è forte al tal punto che il dato digitale, non assume un valore ben definito, ma viene ceduto senza la stesura di un contratto formale tra le parti che ne stabilisca il prezzo. Ciò converge verso un ecosistema dominato da poche multinazionali, aventi un alto grado di integrazione verticale in tutte le fasi, che coesistono con tante piccole imprese specializzate (AGCOM[3], 2018).

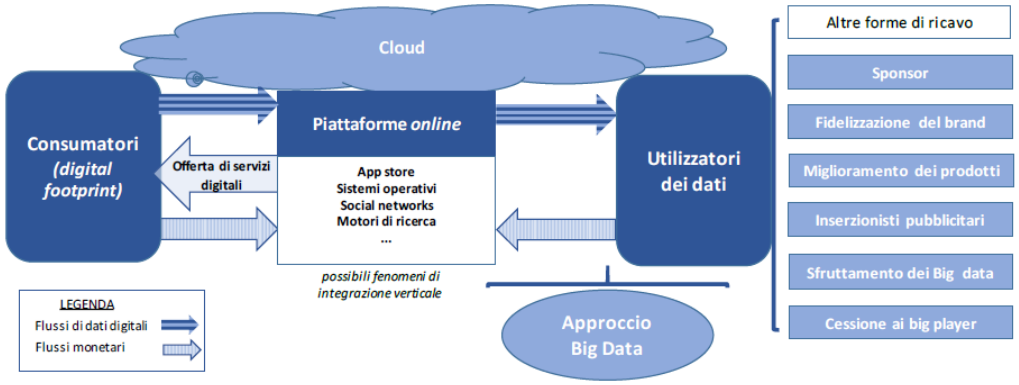

Figura 3: Schema di un mercato a due versanti applicato ai dati digitali
Fonte: (AGCOM, 2018)

Tra gli attori principali all'interno dell'ecosistema dei Big Data è possibile identificare (AGCOM, 2018):

- i fornitori di dati, tipicamente i consumatori.
- le piattaforme.

[3] "Autorità per le Garanzie nelle Comunicazioni"

- gli utilizzatori dei dati (generalmente imprese o enti pubblici ma anche singoli individui) che usano i Big Data per creare valore aggiunto.
- i data brokers (di cui si parlerà ampiamente nel Capitolo 2), cioè organizzazioni o individui specializzati nella raccolta di dati provenienti da varie fonti (sia pubbliche che private) che rivendono poi a terzi per vari scopi.

I Big Data stanno diventando sempre più un asset strategico per aziende e governi. Quindi, comprendere aspetti e peculiarità è essenziale per sfruttarne appieno il grande potenziale.

1.1 Big Data: definizione e le cinque "V"

Non esiste una definizione univoca e universalmente accettata per i Big Data, ma in generale si fa riferimento a una vasta mole di dati digitali che le organizzazioni e i governi acquisiscono, collezionano e analizzano (Big Data Analytics) al fine di creare valore aggiunto (Bagnoli, V., 2017).

Tuttavia, a seconda del focus sul quale si focalizza la descrizione del fenomeno, è possibile suddividere le definizioni esistenti in quattro gruppi (De Mauro et al. 2016):

- il primo gruppo pone il focus sulle loro caratteristiche. Una delle definizioni più diffuse riguarda le cosiddette "5 V" dei Big Data (Bagnoli, V., 2017). Inizialmente le V erano tre ovvero, "Volume, Velocità e Varietà" (Laney, 2001). In seguito, furono aggiunte le altre due "V", ovvero "Veracità" (Schroeck et al., 2012) e "Valore" (Dijcks, 2013). Si è arrivati fino a "42 V" (Shafer, T., 2017). Le cinque "V" tengono conto dell'essenza dinamica dei Big Data. L'International Data Corporation (2011) definisce i Big Data come "una nuova generazione di tecnologie e di architetture progettate per estrarre valore economico da volumi molto grandi di dati, volumi molto grandi di un'ampia varietà di dati, consentendo l'acquisizione, l'analisi e/o la scoperta ad alta velocità". Tale definizione ha fornito le basi per giungere a quella che è considerata ad oggi una delle definizioni di Big Data più semplici ed oggettive, nello specifico i Big Data sono definiti come un "asset informativo

caratterizzato da un volume, una velocità e una varietà tali da richiedere tecnologie e metodi analitici specifici per la sua trasformazione in valore" (De Mauro et al. 2016);

- nel secondo gruppo l'enfasi ricade sulle tecnologie necessarie per elaborare grandi quantità di dati poiché "le tecniche tradizionali per lavorare con i dati non sono più sufficienti "ma "possono essere elaborati solo con difficoltà utilizzando gli strumenti di gestione dei database tradizionali "(Loukides, M., 2010). Da ciò nasce l'esigenza di una "architettura scalabile per un'archiviazione, manipolazione e analisi efficiente" (National Institute of Standards and Technology, 2014);
- alcune definizioni trattano i Big Data in termini di superamento di determinate soglie: ad esempio, quando la quantità di dati supera la capacità di elaborazione dei sistemi di database convenzionali, rendendo indispensabile l'adozione di sistemi di elaborazione alternativi (Dumbill, 2013);
- nel quarto e ultimo gruppo le definizioni si concentrano sull'impatto dei Big Data sulla società. Una popolare definizione definisce i Big Data come "un fenomeno culturale, tecnologico e accademico basato su tre elementi: tecnologia, analisi e mitologia ("la percezione che i grandi dataset portino con sé verità e accuratezza") (Boyd & Crawford, 2012);

Per comprendere meglio la complessità e il potenziale dei Big Data si può partire dalle definizioni appartenenti al primo gruppo analizzandone le caratteristiche intrinseche (le cosiddette "cinque V") (AGCOM, 2018):

- **Volume:** è senza dubbio la caratteristica che più naturalmente si può collegare ai Big Data. Si fa riferimento alla grande quantità di dati oggi disponibili che compongono quella che è definita la **datasfera.** Non si conosce esattamente l'ammontare del fenomeno, ma gli studi ad oggi disponibili parlano di una dinamica di crescita esponenziale. Secondo un IDC ("International Data Corporation") forecast del 2021 nel 2025 si prevede di raggiungere i 181 ZB di dati prodotti (circa il 448 % in più rispetto ai 33 ZB del 2018). (STATISTA,

2021). Tale crescita può essere associata ad un nuovo paradigma per cui le aziende stanno passando dai database tradizionali ai cloud (pubblici e privati) per le loro operazioni di elaborazione dati. Si stima, infatti, che nel 2025 il 49% dei dati mondiali sarà conservato in cloud pubblici (Reinsel et al. 2018);

- **Varietà:** si riferisce all'eterogeneità sia delle fonti sorgenti dei dati (social media, sensori IoT, transazioni economiche, ricerche sul web ecc.), sia dei diversi formati con cui le informazioni vengono raccolte. Si possono avere **dati strutturati**, cioè, organizzati in strutture composte da righe e colonne. Tali dati vengono processati agevolmente con tecniche consolidate che si rifanno ai sistemi di database relazionali (Customer Relationship Management, RDBMS[4], fogli di calcolo ecc.). Tuttavia, in campo Big Data, i dati si presentano perlopiù in forma **non strutturata** poiché, non hanno una struttura ben definita (foto, e-mail, sensori, social media ecc.). Dunque, non è possibile utilizzare le tecniche classiche (database relazionali) ma è necessario adoperare tecniche di elaborazione più sofisticate (AGCOM, 2018);

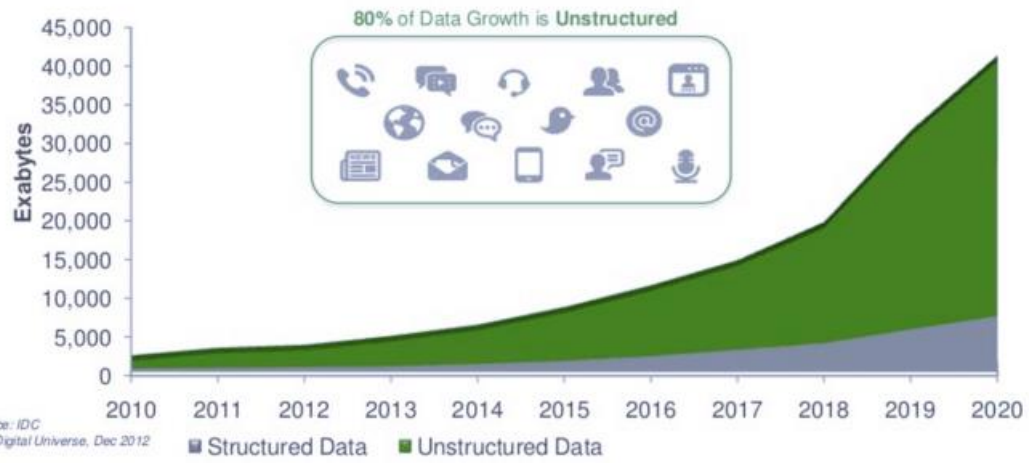

Figura 4: Trend di crescita dei dati non strutturati (1 Exabytes= 10^{-3} ZB)
Fonte: (IDC The Digital Universe. The Massive Growth in Unstructured Data, 2012)

[4] "Relational Database Management System". Ovvero programmi per creare e gestire database relazionali. Tra i più famosi ricordiamo Microsoft SQL Server e Oracle Database.
Fonte: (https://en.wikipedia.org/wiki/Relational_database)

- **Velocità:** si riferisce non solo alla velocità di circolazione dei dati ma anche alla velocità con la quale vengono raccolti e processati. I dati sono risorse deteriorabili nel tempo e per sfruttarne appieno il valore è cruciale una rapida e tempestiva elaborazione (AGCOM, 2018);

Figura 5: Dati generati nel mondo in 1 minuto nel 2023.
Fonte: (https://www.domo.com/learn/infographic/data-never-sleeps-11)

- **Valore:** rappresenta un punto fondamentale, in quanto l'elaborazione dei Big Data ha come scopo quello di estrarre valore dai dati e ottenere *insight* mirati a guidare le decisioni strategiche delle imprese. Tuttavia, non si tratta solo di aumentare il profitto delle aziende, ma anche il benessere della società nel suo insieme. Inoltre, è importante sottolineare che i dati hanno un valore di "primo utilizzo" (ad esempio, un sito di e-commerce può acquisire dati del comportamento del cliente in tempo reale e offrire un'offerta mirata) e un "valore opzionale o secondario", cioè derivante dal riutilizzo dei dati per scopi diversi da quelli del loro utilizzo primario (ad esempio, Google nel 2008 utilizzò i dati raccolti dalle ricerche degli utenti sull'influenza stagionale per mettere a punto un algoritmo che prevedesse e mappasse in tempo reale la diffusione dell'influenza (Google Flu Trends Estimates, 2008). Allo stesso

modo, durante la pandemia di Covid-19, i dati presenti nei registri ospedalieri furono utilizzati per la ricerca e lo sviluppo di nuovi farmaci contro la malattia.

- **Veridicità:** riguarda la qualità e l'affidabilità dei dati raccolti. È essenziale che le organizzazioni lavorino con dati affidabili in modo tale che le analisi portino a *insight* corretti. La veridicità si estende anche alla gestione delle incertezze e delle ambiguità dei dati (AGCOM, 2018);

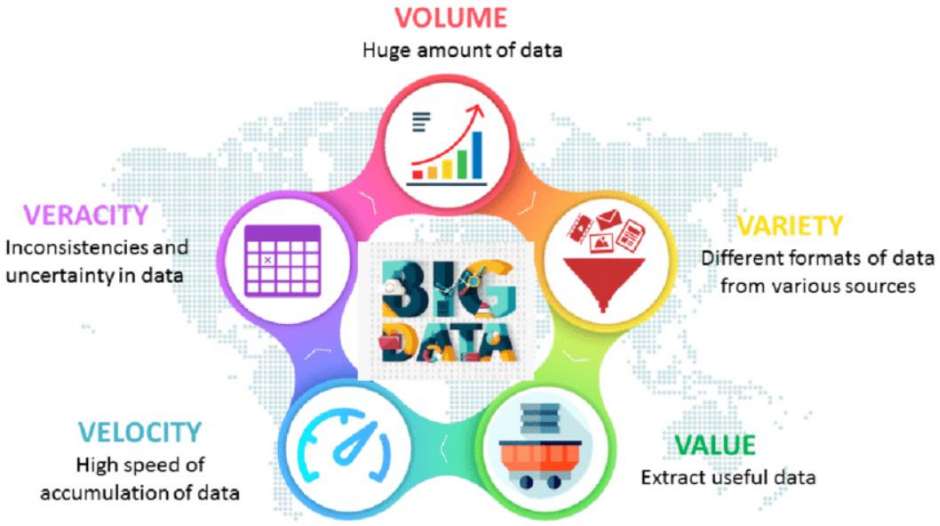

Figura 6: Le cinque "V" dei Big Data.
Fonte: (https://www.researchgate.net/figure/Figura-6-Le-5-V-dei-Big-Data_fig1_342611818)

1.2 La Data Value Chain

La **catena del valore** è un concetto generale che può essere esteso a vari ambiti e la sua definizione dipende dal contesto al quale facciamo riferimento. Nel campo del Business Management la catena del valore è stata introdotta da Porter (1985) e consiste "nell'analisi di tutte le attività e delle loro interazioni, nell'identificare le fonti di un potenziale vantaggio competitivo e organizzare i processi che organizzare i processi che contribuiscono direttamente o indirettamente alla creazione di valore aggiunto per l'organizzazione" (Faroukhi et al. 2020). Le attività si possono suddividere in "**attività primarie"** (cioè, le attività che l'azienda svolge per produrre e vendere il proprio prodotto e che di conseguenza hanno impatto diretto sulla creazione di valore) e

"**attività di supporto"** che non creano direttamente valore, ma sono essenziali per il funzionamento delle attività primarie (Porter, M., 1985)

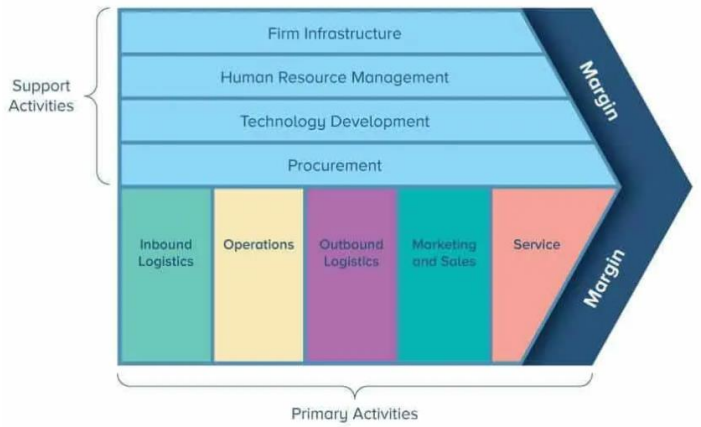

Figura 7: Catena del Valore di Porter (1985)
Fonte: (https://www.smartsheet.com/value-chain-model)

La Catena del Valore di Porter (1985) si concentrava principalmente sui processi fisici e logistici; tuttavia, l'avanzare delle nuove tecnologie ha portato alla necessità di estendere questa teoria anche al mondo digitale. Dunque, la "**Data Value Chain**" non è altro che l'applicazione della teoria di Porter al mondo digitale ed è uno strumento utile per identificare le varie fasi attraverso le quali passano i dati. Tali fasi vanno dall' **acquisizione** del dato all'**immagazzinamento** fino ad arrivare all'**utilizzo** del dato, che rappresenta l'effettiva creazione di valore aggiunto. Il dato si muove tra queste fasi tra loro interconnesse che ne accrescono il valore progressivamente durante il processo (AGCOM, 2018): tale processo può essere definito come "ciclo di vita del dato" (FTC[5] REPORT, 2016).

[5] "Federal Trade Commission". E' un'agenzia governativa statunitense che ha il compito di "promuovere la tutela dei consumatori, l'eliminazione e la prevenzione di pratiche commerciali anticoncorrenziali". Fonte: https://it.wikipedia.org/wiki/Federal_Trade_Commission

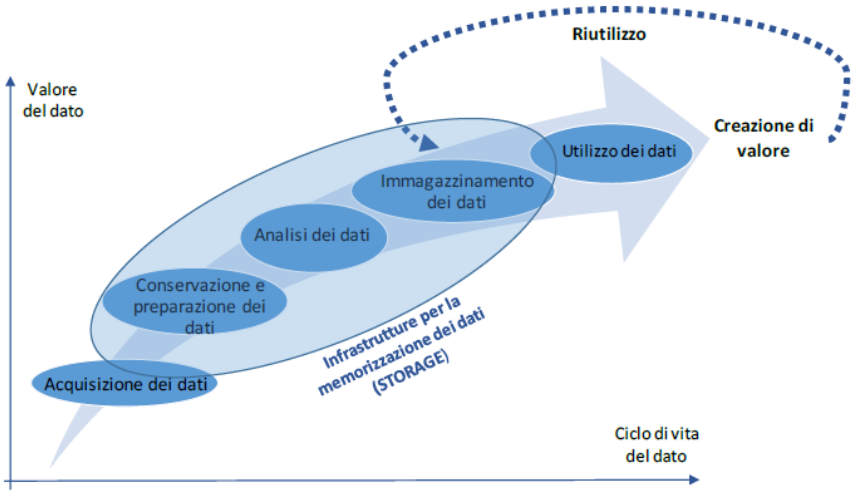

Figura 8: La Value Chain nei big data
Fonte: (AGCOM, 2018)

- **Acquisizione dei dati**: il dato viene acquisito in diversi modi. Si definisce **acquisizione diretta** quando il contatto avviene direttamente col soggetto che fornisce il dato. Alcuni dati vengono forniti volontariamente dai consumatori con un "click", ad esempio quando ci si registra o si effettua il login ad un sito. Tuttavia, è possibile che i dati siano collezionati tracciando i consumatori online tramite *cookies di tracciamento, history sniffing* (la pratica di tracciare lo storico dei siti visitati dall'utente) o *web scraping* (degli appositi script che estraggono dati dai siti web). Oltre all'*acquisizione diretta* i dati possono essere acquistati **indirettamente** da terze parti, ovvero dai **data broker** (vedi Cap. 2) (Bourreau et al. 2017).
- **Conservazione e preparazione dei dati**: è la fase in cui i *raw data* (dati grezzi) iniziano a trasformarsi in informazione. Affinché questo processo avvenga correttamente è necessario dotarsi di un'architettura che sia adeguata in materia di sistemi di elaborazione e software. Inoltre, per affrontare le sfide legate a "velocità, varietà e volume" dei dati è fondamentale che le tecnologie e le competenze siano estremamente flessibili. Un nuovo paradigma riguarda i cosiddetti *data lake* (cioè, database centralizzati in cui i dati vengono archiviati in forma grezza e nel loro formato originale) in contrapposizione ai tradizionali *data silos* e *data warehouse* aziendali che archiviano i dati in modo strutturato (tipicamente organizzati in

tabelle o secondo uno schema gerarchico) e godono di scarsa *scalabilità* e *flessibilità*.

Il vantaggio offerto dai data lake è dunque la possibilità di archiviare una vasta gamma di dati strutturati e non, come ad esempio: documenti, e-mail, immagini, video, file multimediali ecc. Tuttavia, quando non è necessario che i dati comunichino tra loro o per la comune reportistica quotidiana, i sistemi tradizionali risultano adeguati (AGCOM, 2018).

- **Analisi dei dati**: è il primo step della catena del valore in cui il dato passa da "semplice informazione a conoscenza del fenomeno che si sta analizzando" (AGCOM, 2018).

  L'analisi dei dati "si occupa di rendere i dati grezzi acquisiti utilizzabili nel *decision making* e nell'uso specifico del dominio" (Cavanillas et al. 2016). Essa comprende attività come "esplorazione, trasformazione e modellazione con lo scopo di evidenziare i dati rilevanti e portare alla luce informazioni nascoste che hanno un alto potenziale strategico per l'azienda" (Cavanillas et al. 2016). Degli esempi sono: il business intelligence, data mining e machine learning.
- **Immagazzinamento dei dati**: ciascuna organizzazione progetta l'immagazzinamento dei propri dati in modo diverso; tuttavia, è importante che le infrastrutture di memorizzazione siano facilmente *scalabili*, considerando la crescita inarrestabile del volume dei dati. Avere sistemi facilmente "*scalabili*" è indispensabile per avere un accesso rapido ai dati e velocizzare il *decision making.* Difatti, il trend attuale è quello di utilizzare dei *database distribuiti* in cui i dati sono sparsi sulle memorie dei diversi *nodi* (computer) che compongono la rete di un'organizzazione. Tali nodi non sono necessariamente vicini tra loro, ma possono anche trovarsi a grandi distanze fisiche come, ad esempio, la rete di una grande multinazionale con sedi in tutto il mondo (AGCOM, 2018). L'utilizzo delle tradizionali tecnologie di sistemi di database (RDBMS), caratterizzate tipicamente da un unico componente hardware dedicato all'immagazzinamento, non appaiono più compatibili con le tendenze attuali (Cavanillas et al. 2016).

- **Utilizzo dei dati**: concerne l'ultima fase della *catena del valore*. I dati, dunque, possono essere **utilizzati** per molteplici scopi tra cui: supporto ai processi decisionali, personalizzazione di prodotti e servizi, prezzi personalizzati, targeted advertising, gestione del rischio e prevenzione delle frodi, pricing dinamico ecc. In alternativa i dati possono essere **riutilizzati** per scopi secondari (vedi par.1.1). Il paragrafo 1.3 si propone di fornire una panoramica generale dei possibili utilizzi dei Big Data.

1.3 Utilizzo dei Big Data

1.3.1 Discriminazione di prezzo

La **discriminazione di prezzo** avviene quando per la medesima unità di un bene o servizio si applicano prezzi diversi a differenti consumatori, basandosi su caratteristiche come fascia d'età, comportamento d'acquisto, posizione geografica, disponibilità a pagare, fedeltà al brand ecc. E' importante evidenziare che la "differenza di prezzo non riflette una differenza di costi" (Bourreau et al. 2017).
Si distinguono tre forme di discriminazione di prezzo (Pigou, 1920):

- **Primo grado**: si stabilisce un prezzo diverso per ogni unità di un prodotto e tale prezzo è esattamente uguale alla massima **disponibilità a pagare** del cliente, così facendo l'impresa cattura *l'intero surplus del consumatore.* È necessario conoscere *perfettamente* le preferenze dei consumatori (Varian H. 1987).
- **Secondo grado**: si verifica quando l'impresa fissa prezzi differenti in base al numero di unità acquistate. Un esempio classico sono sconti e premi se si acquistano grandi quantità del bene (Varian, H., 1987).

- **Terzo grado**: detta anche *group pricing*, l'impresa non osserva *perfettamente* l'eterogeneità dei consumatori ma riesce a osservare dei gruppi a cui applicare prezzi differenti. Così facendo si è in grado di "discriminare tra i gruppi ma non all'interno del gruppo" (Bourreau et al. 2017). Un esempio classico sono gli sconti per gli studenti universitari o i prezzi dei voli differenti a seconda del posizionamento geografico dei clienti (Varian, H., 1987).

L'uso dei Big Data facilita le strategie di discriminazione di prezzo operate dalle imprese, che utilizzano i dati raccolti online o acquistati da terze parti (vedi Data Broker nel Cap.2) per ottenere informazioni sulle preferenze dei consumatori e sulla loro **WTP** (Willingness To Pay). È così possibile tracciare le preferenze dei consumatori con un livello di dettaglio molto fine e personalizzato. Gli effetti della **price discrimination** sono ben visibili anche sul mercato dei prodotti digitali, dal momento che le imprese potrebbero aver incentivo ad abbassare la qualità dei prodotti meno costosi, in modo da attrarre consumatori con un'alta WTP e che dunque sono disposti a pagare tanto per prodotti di alta qualità. L'idea è di offrire "differenti qualità a differenti prezzi". Inoltre, la price discrimination può ridurre la fiducia nei mercati online poiché può portare i consumatori non solo a percepire il mercato come poco trasparente ma anche minare la fiducia nei confronti degli altri operatori del mercato online (Bourreau et al., 2017).
Sebbene la discriminazione dei prezzi sia potenzialmente applicabile a diversi mercati e settori, ad oggi risulta ancora difficile trovare in letteratura delle evidenze sull'uso dei dati per questo fine (OECD, 2018). I motivi sono molteplici: è possibile che molte aziende siano restie ad adottare questa strategia per paura di perdere reputazione o generare una reazione negativa da parte dei consumatori (Council of Economic Advisers, 2015), oppure è possibile che le aziende implementino strategie di discriminazione di prezzo, ma in modi non trasparenti. In ogni caso, rilevare il pricing personalizzato è un compito complesso perché le "tecniche di personalizzazione online sono diventate molto più avanzate e difficili da catturare/misurare" (European

Commission, 2018). Di seguito alcune evidenze sull'impiego dei dati ai fini di attuare pratiche di discriminazione dei prezzi:

- Lo studio di Hannak et al. (2014) dal titolo *Measuring Price Discrimination and Steering on E-commerce Web Sites* ha analizzato 16 siti web di e-commerce, riscontrando la presenza di forme di discriminazione dei prezzi in 9 di essi. Dallo studio emerge che i dati raccolti dai consumatori per attuare la discriminazione erano perlopiù: dati sull'iscrizione, cronologia dei click e storico degli acquisti passati;
- Nel 2016 si è scoperto che la piattaforma online Coupons.com utilizzava dati sul comportamento dei suoi utenti (preferenze d'acquisto, carte fedeltà e tipi di prodotti cercati) per targetizzare gli utenti e proporre prezzi personalizzati (Ezrachi, A et al., 2016);
- Nel 2018 Uber ha utilizzato dati e algoritmi di machine learning per implementare un sistema di prezzatura che sembra rientrare nell'ambito dei prezzi personalizzati. Sembra che l'azienda abbia attuato una strategia di "discriminazione di prezzo di terzo grado", suddividendo i consumatori in gruppi diversi in base al percorso scelto dal cliente e all'orario della corsa, per poter addebitare tariffe più alte ai gruppi di consumatori che si spostano tra quartieri benestanti (OECD, 2018). Il responsabile del prodotto in un'intervista a Bloomberg ha infatti dichiarato "l'azienda applica algoritmi di machine learning per stimare quanto i gruppi di clienti sono disposti a pagare per una corsa […]" (Newcomer, E., 2017).

1.3.2 Target advertising

Un altro impiego dei Big Data è la pratica del **target advertising** (pubblicità mirata), cioè l'utilizzo dei dati sul comportamento online per offrire una pubblicità mirata a gruppi specifici di consumatori, aumentando così la probabilità che l'annuncio sia per loro rilevante e venga di fatto tramutato in un acquisto.

Esistono due forme principali di target advertising: **search advertising** e **non-search advertising.**

Il **search advertising** è basato sulle ricerche dei consumatori (ad esempio inserzionisti su Instagram possono proporre annunci sulla base delle ricerche degli utenti; ad esempio, se un utente effettua tante ricerche sulla moda, Instagram proporrà annunci di abbigliamento).

Il **non-search advertising** riguarda i banner pubblicitari o annunci video che appaiono su siti web, app o piattaforme di social media (ad esempio gli annunci che compaiono su You Tube prima o dopo la riproduzione del video) (FTC REPORT, 2016).

Secondo IAB Europe, nel 2015 il 47% del mercato dell'online advertising era rappresentato dal "search advertising", mentre il restante 53% dal "non-search advertising".

La pubblicità mirata è quindi una pratica molto diffusa, come sottolineato dal lavoro di Carrascosa et al. (2015), che ha dimostrato che l'88% dei "profili utente simulati" (per condurre l'esperimento sono stati creati appositamente 72 "profili utente simulati", ciascuno con caratteristiche ben precise come: interessi personali, abitudini di navigazione ecc.) riceveva pubblicità mirata in base agli interessi rilevati. Inoltre, lo studio ha rilevato come gli inserzionisti, per personalizzare gli annunci, usino anche informazioni sensibili relative a salute, orientamento politico, religioso ecc. mostrando una notevole precisione nella personalizzazione degli annunci.

Molte evidenze empiriche sull'utilizzo dei dati per implementare queste pratiche, si trovano in una particolare forma di targeted advertising, il **retargeting**, che consiste nel mostrare annunci personalizzati (basati sulle interazioni con un sito web già visitato in precedenza dall'utente), mentre naviga su un altro sito (Bourreau et al., 2017).

Il **target advertising** presenta dei benefici per i consumatori legati soprattutto ad un miglioramento dell'esperienza utente. Questo succede perché gli annunci si allineano meglio con gli interessi dell'utente facilitando così l'accesso ai prodotti da loro desiderati e riducendo anche i *searching cost* degli utenti che così possono prendere decisioni d'acquisto più rapidamente.

Tuttavia, il target advertising porta con sé dei rischi notevoli come problemi di privacy (i consumatori possono sentirsi a disagio per il fatto che i loro dati personali vengono tracciati con relativa facilità e utilizzati, senza il loro consenso, per fissare prezzi personalizzati) o rischi di manipolazione (per cui i consumatori, bombardati dagli annunci, vengono indotti in modo ingannevole verso acquisti costosi o non necessari) (FTC REPORT, 2016). Un altro aspetto importante da tenere in considerazione è che la pubblicità mirata può precludere ai consumatori l'esposizione ad una vasta gamma di prodotti o servizi, limitando di fatto le loro scelte a un numero ristretto di alternative.

1.3.3 Ulteriori utilizzi dei Big Data

Di seguito, un elenco di altri utilizzi dei Big Data:

- **supporto ai processi decisionali**: collegando i dati raccolti alle azioni intraprese dall'organizzazione è possibile migliorare: i processi decisionali interni, l'organizzazione del personale, le procedure interne, prodotti e servizi esistenti, nonché è possibile ottenere insight utili per migliorare i KPI (AGCOM, 2018);
- **miglioramento dell'efficienza produttiva**: la raccolta e l'analisi dei dati relativi ai processi interni permette di individuare i punti di scarsa produttività e intervenire per migliorare quest'ultima. Ciò può portare diversi benefici alle aziende, ad esempio una riduzione dei costi di produzione (AGCM[6] et al., 2020);
- per **offrire prodotti e servizi innovativi** che diversamente non potrebbero essere realizzati. Un esempio sono i servizi che informano gli utenti delle condizioni del traffico sulle arterie stradali, i quali sono realizzati attraverso la raccolta e l'analisi dei dati di posizione e di spostamento di milioni di singoli utenti (AGCM et al., 2020);
- in **ambito sanitario** sono stati creati database ad accesso libero, contenenti dati clinici di pazienti in forma anonima, la cui analisi può aiutare gli scienziati a

[6] "Autorità Garante della Concorrenza e del Mercato"

estrarre “nuova conoscenza in maniera automatizzata su una determinata patologia” (AGCM et al., 2020);

- **personalizzazione di prodotti e servizi**: ad esempio le piattaforme che distribuiscono contenuti digitali o di e-commerce analizzano i Big Data per proporre ai propri utenti beni e servizi in linea con le preferenze individuali. Alcune piattaforme online acquisiscono dati personali e relativi alle abitudini dei consumatori per implementare forme di “search discrimination”, ovvero personalizzano la visualizzazione dei risultati di ricerca online (AGCM et al., 2020). Non solo le piattaforme digitali, ma anche le aziende operanti in settori più tradizionali (come il manifatturiero), possono trarre beneficio dall’analisi dei Big Data. Quest’ultime possono utilizzare i dati sui consumatori e sui trend di mercato per sviluppare prodotti che si adattano meglio alle esigenze dei clienti (AGCM et al., 2020);
- sta crescendo inoltre l’interesse degli istituti di credito e delle compagnie assicurative riguardo l’impiego dei Big Data per la **prevenzione delle frodi** e la **mitigazione del rischio creditizio/assicurativo**. Tuttavia, sebbene gli operatori del settore riconoscano la potenzialità dei Big Data, l’approccio rimane “prudenziale”, in quanto allo stato attuale, non vi è certezza di un ritorno economico a fronte degli ingenti investimenti da fare (AGCM et al., 2020). Esistono delle evidenze di compagnie assicurative che hanno investito per diventare “data-driven company”, ad esempio Cattolica Assicurazioni dal 2018 ha deciso di investire sui big data e sul data management focalizzandosi anche sulla prevenzione delle frodi. In particolare, sono stati realizzati modelli di machine learning per “l’individuazione e il riconoscimento di pattern potenzialmente fraudolenti” (Politini, S., 2020).
- la nuova frontiera, spinta dalla diffusione dell’Internet of Things e dell’Intelligenza Artificiale, è quella delle **smart city** e delle **smart manifacturing**. In quest’ultimo caso, l’Intelligenza Artificiale analizzando i dati prodotti dai sensori di monitoraggio distribuiti negli impianti, applica modelli previsionali che

permettono di stabilire in anticipo le probabilità di anomalie e fermi macchina ("**manutenzione predittiva**") (Zanotti, L., 2024).

Secondo un report di Rockwell Automation (2024), gli investimenti in smart manifacturing sono aumentati del 30% rispetto al 2023.

Capitolo 2: I data broker: la vendita a terzi dei dati

2.1 I Data Broker: introduzione

Dopo aver descritto brevemente le principali caratteristiche e i possibili utilizzi dei Big Data, il seguente capitolo porrà l'accento sul ruolo dei **data broker** nell'ecosistema dei dati. A causa della natura variegata e della vasta portata del fenomeno, ad esempio i diversi modelli di business o i diversi modi in cui viene estratto il valore dai dati, non esiste una definizione pienamente esaustiva e uniforme per tutti dei data broker. Secondo una definizione della US Federal Trade Commission (FTC) i **data broker** sono "società che raccolgono informazioni, comprese quelle personali sui consumatori, da un'ampia varietà di fonti per rivendere tali informazioni ai propri clienti per vari scopi, tra cui la verifica dell'identità di un individuo, la differenziazione dei record, la commercializzazione di prodotti e la prevenzione di frodi finanziarie" (Federal Trade Commission, 2012). Tali informazioni possono includere nome e cognome dell'individuo, indirizzo di casa, educazione scolastica, livello di reddito, stato civile, interessi politici e sociali, gusti musicali, dati sui viaggi effettuati, informazioni finanziarie, transazioni effettuate con la carta di credito, stato di salute ecc.

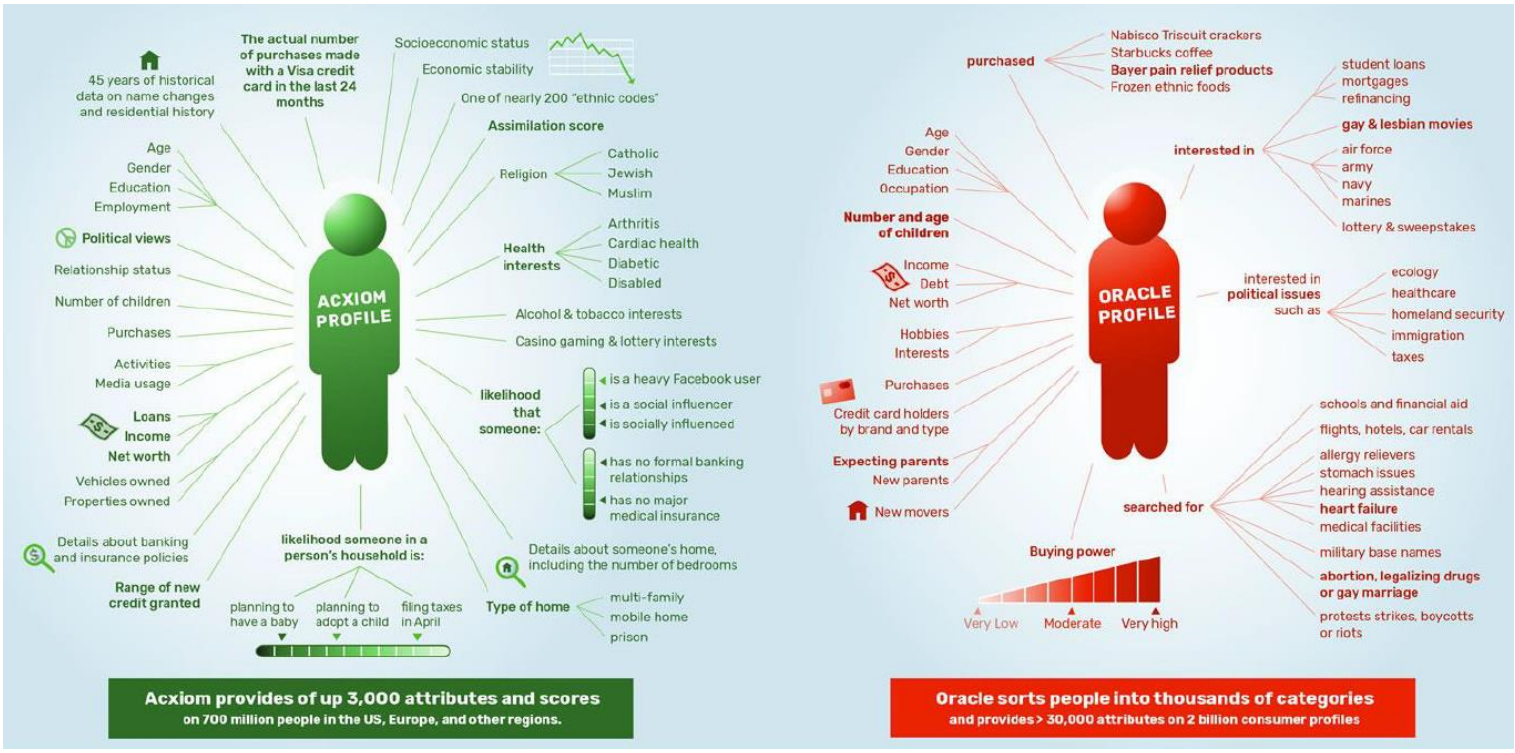

Figura 9: Dati sui consumatori forniti da due importanti data broker, Acxiom e Oracle.
Fonte: (Corporate Surveillance In Everyday Life, 2017)

I data broker, dunque, dispongono di un'enorme quantità di informazioni; tuttavia, la portata del fenomeno resta tutt'oggi in grande inesplorata e sconosciuta.

		Credit Reporting Agencies	
Experian	has credit data on	918 million	people
	marketing data on	700 million	people
	„insights" on	2.3 billion	people
Equifax	has data on	820 million	people
		1 billion	devices
TransUnion	has data on	1 billion	people
		Consumer Data Brokers	
Acxiom	has data on	700 million	people
		1 billion	cookies and mobile devices
	it manages	3.7 billion	consumer profiles for clients
Oracle	has data on	1 billion	mobile users
		1.9 billion	website visitors
	provides access to	5 billion	"unique" consumer IDs

Figura 10: Quantità di dati sui consumatori posseduti da alcune data companies.
Fonte: (Corporate Surveillance In Everyday Life, 2017)

Questi dati hanno un grande mercato e possono essere acquistati dalle aziende per vari scopi, ad esempio nell'ambito delle piattaforme digitali, è possibile utilizzare i dati raccolti dalle interazioni quotidiane degli utenti, come ricerche sul web o acquisti online, per tracciare il comportamento e i gusti dei consumatori. Imprese come Google e Facebook possono così utilizzare i dati forniti col consenso degli utenti (della privacy si discuterà in modo approfondito nel Cap. 4) per dare ad altri soggetti la possibilità di pubblicare annunci mirati sulle loro piattaforme (vedi Cap.1 par.1.3.2). Altri possibili acquirenti sono le compagnie assicurative che acquistano i dati per migliorare la valutazione del rischio di un potenziale cliente o gli istituti di credito che possono acquistare i dati sulla storia creditizia passata per valutare correttamente il merito creditizio di un potenziale richiedente prestito (ACCC[7], 2023). I data broker, dunque, sono "compagnie che collezionano dati sui consumatori, da fonti online e offline, combinano i dati creando dei profili sui consumatori che vendono o condividono ad altre compagnie o ad altri data broker" (Federal Trade Commission, 2014). Tuttavia, il termine "data broker" è più comune negli Stati Uniti che in Europa, infatti nel vecchio continente si è soliti riferirsi ai data broker come "information resellers," "data suppliers,", "data aggregators", "consumer data analytics" (Rieke, A. et al., 2016).

[7] "Australian Competition and Consumer Commission's". Ovvero l'autorità regolatore della concorrenza in Australia. Fonte: https://en.wikipedia.org/wiki/Australian_Competition_and_Consumer_Commission

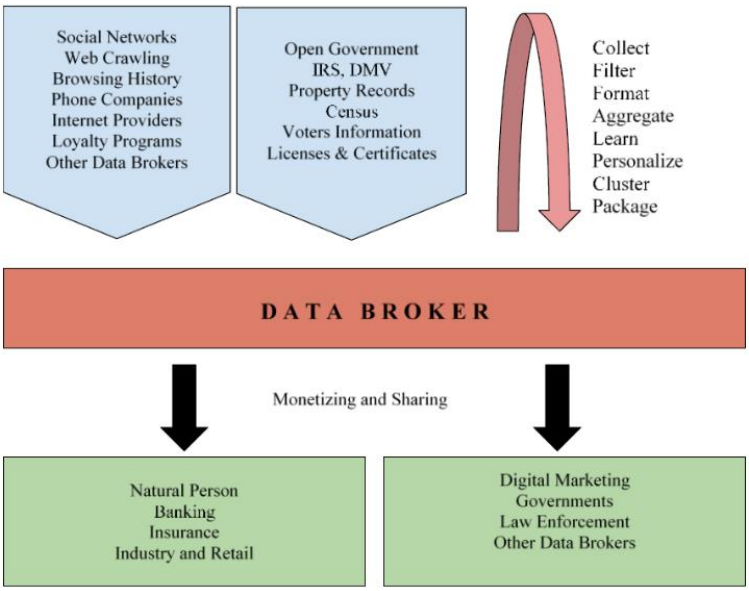

Figura 11: Flusso di informazioni da/verso un data broker.
Fonte: (Birckan, G. et al., 2020)

Non tutte le aziende che raccolgono e vendono dati rientrano in questa definizione; infatti, per essere considerato un data broker, non è sufficiente soltanto raccogliere dati e potenzialmente ottenere delle entrate dalla vendita, ma questa attività deve essere anche la "principale fonte di ricavi dell'impresa". Ad esempio, Disney, in base alla definizione data in precedenza, non è un data broker perché pur condividendo dati sui consumatori con altre società controllate, tale attività non costituisce la principale fonte di profitto. In modo analogo, Netflix, raccoglie dati sulle preferenze e le abitudini di visione degli utenti, ma non vende questi dati a terzi soggetti. Google e Facebook, sebbene raccolgano una grande quantità di informazioni personali dagli utenti, non vendono i dati ad altre aziende, ma li utilizzano per permettere agli inserzionisti di "indirizzare gli annunci sulla loro piattaforma agli utenti più propensi a rispondere" oppure per migliorare l'esperienza utente tramite lo sviluppo di nuovi prodotti e servizi; quindi, in base a quanto detto in precedenza, non possono essere considerati data broker (Rieke, A. et al., 2016).

È opportuno precisare che a differenza della vendita, nel caso della condivisione l'unica aspettativa non è quella di ricevere un corrispettivo in denaro, ma anche altri benefici quali scambi di informazioni o collaborazioni strategiche. (Reviglio, U., 2022).

A seconda della modalità di acquisizione dei dati è possibile dividere i data broker in due categorie:

- **First-party data brokers**: ovvero imprese che raccolgono dati direttamente dai consumatori, senza intermediazione di terze parti. Ad esempio, possono provenire da interazioni sui siti web, transazioni e-commerce, social media, utilizzo di app, sondaggi o iscrizioni a newsletter. Poiché sono raccolti direttamente dalla fonte, tali dati vengono considerati più accurati e significativi per l'azienda che li possiede.
- **Third-party data brokers:** le imprese in questo caso raccolgono i dati sui consumatori da una varietà di terze fonti, ovvero senza interazione diretta col consumatore. (ACCC, 2023). La maggior parte dei data broker ad oggi conosciuti ricadono in questa categoria (Christl, W., 2017).

2.2 L'ecosistema dei data broker

Il brokeraggio dei dati non è una pratica nuova, ma precede l'avvento dell'era digitale. Negli anni Settanta, ad esempio, la data company Claritas, presentò il cosiddetto "sistema di segmentazione dello stile di vita" (chiamato PRIZM) per ottenere degli insight sulle abitudini d'acquisto dei consumatori Americani che potessero essere d'aiuto agli operatori del marketing. I **credit bureaus** nascono negli Stati Uniti intorno agli anni Cinquanta con lo scopo di aiutare, fornendo dati precisi sulla storia creditizia del debitore, i creditori nella concessione del finanziamento. Dunque, in origine nascono come dei "data broker locali e di piccole dimensioni". Se da un lato il brokeraggio dei dati e il profiling non sono concetti nuovi, ciò che è cambiato è l'aumento in volume dei dati prodotti e il numero delle fonti. Ad esempio, le transazioni derivanti dall'uso delle carte di credito per acquisti online, l'uso dei "digital wallet" come Paypal, l'uso degli smartphone che generano dati sulla localizzazione, sulle app utilizzate e sui contatti, le ricerche sul web, sono una ricca fonte di dati. (Rieke, A. et al., 2016) Per quanto concerne la dimensione del mercato, nel 2023 i ricavi derivanti dal brokeraggio dei dati sono stati pari a circa 264 miliardi di dollari (Market

Research Future, 2024) e i forecast per il 2024 parlano di una cifra vicina ai 390 miliardi di dollari (Knowldege Sourcing Intelligence, 2024). Ci si aspetta che la crescita del mercato globale del brokeraggio di dati, nel periodo 2023-2032, avvenga con un CAGR del 7.5% (Market Research Future, 2024).

Il Nord America domina il mercato con un market share superiore al 30%. Tale dato può essere ricondotto sia alla presenza di multinazionali nei settori delle tecnologie digitali, salute, finanza e retail che generando tantissimi dati, spingono verso l'alto i servizi di brokeraggio di dati in questi settori, sia all'assenza di una regolamentazione federale sulla privacy che ponga un limite all'acquisizione aggressiva dei dati. tantissimi dati. Tale primato si riscontra ad esempio nel settore dei prodotti per **la mitigazione del rischio**, settore molto sviluppato in Nord America grazie anche alla presenza di imprese leader nella reportistica di credito come Experian ed Equifax. Al secondo posto troviamo l'Europa, in cui il mercato del brokeraggio si concentra più nel settore manifatturiero e quindi troviamo prodotti come report su performance aziendali e analisi di mercato. Infine, abbiamo il mercato dell'Asia- Pacifica, che si stima sarà il mercato che crescerà di più (i.e CAGR più alto rispetto a Nord America ed Europa) nel periodo 2023-2032. (Market Research Future, 2024).

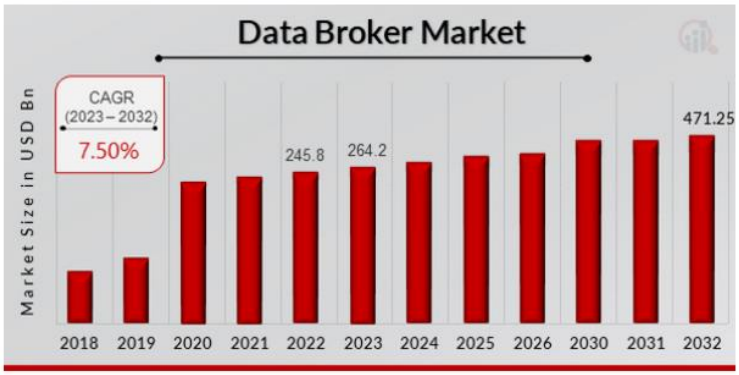

Figura 12: Dimensione mondiale del mercato del brokeraggio dei dati.
Fonte: (https://www.marketresearchfuture.com/reports/data-broker-market-11676)

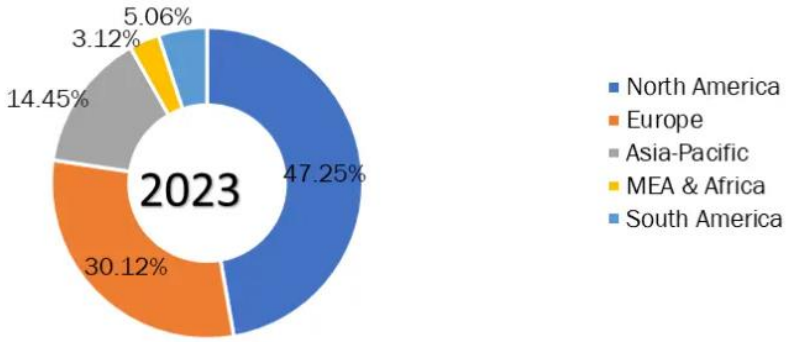

Figura 13: Dimensione del brokeraggio dei dati nel settore assicurativo.
Fonte: (https://www.maximizemarketresearch.com/market-report/insurance-brokerage-market/215727/)

Si stimano circa 4000 di data broker company operanti in tutto il mondo. Tali compagnie possiedono dati su oltre 500 milioni di consumatori in tutto il mondo, acquistati da aziende private e pubbliche e da istituzioni finanziarie. Si stima che l'87% delle aziende in tutto il mondo si rivolga ai data broker per le campagne di marketing mirate (WORLDMETRICS.ORG REPORT, 2024).

Tra i principali data broker esistenti ricordiamo: Experian, Equifax, TransUnion, Acxiom, Oracle, CoreLogic, LexisNexis, Ilion, Verisk, Nielsen le quali vendono "prodotti su misura per diversi scopi e in diversi settori". Tra i settori in cui si inseriscono i data broker menzioniamo: advertising, prevenzione delle frodi, salute, educazione e law enforcement. (Rieke, A. et al., 2016).

Figura 14: Alcuni data broker presenti nel mercato.
Fonte: (https://urna.winstonsmith.org/materiali/2021/atti/ep2021se_24_reviglio_data-brokers.pdf)

Experian, è una multinazionale con sede in 30 paesi nel mondo, che opera nel settore del **credit reporting.** Si occupa dunque di fornire report di credito e prodotti per la gestione del rischio sia alle imprese che ai consumatori. In aggiunta, offre anche "data analytics" per il marketing e "data analytics tool"**,** che vengono poi utilizzati dalle imprese per ottenere insights sul proprio business e prendere decisioni basate sui dati. Come descritto in precedenza, il suo core business è rappresentato dal credit reporting che costituisce il 52 % dei loro ricavi totali, mentre i servizi per la verifica dell'identità rappresentano il 27 % dei ricavi, e infine i data analytics per il decision making e gli analytics per il marketing costituiscono il 21%. I ricavi stimati per il 2024 sono pari a circa 7 miliardi di dollari, di cui il 66 % in Nord America (Experian official website, 2024). Si stima che Experian abbia report di credito su 918 milioni di persone (Experian Annual Report, 2016) e processi ogni mese oltre 2 miliardi di record. Inoltre, possiede "8 miliardi di combinazioni di nomi e indirizzi" che converte in insight utili (Sherman, 2021).

Oltre ai report di credito, come detto in precedenza, offre servizi di verifica dell'identità e deterrenza delle frodi. Per fornire il servizio viene utilizzata una piattaforma chiamata "Precise ID" che si basa su dati provenienti da "immatricolazioni di auto, numeri di previdenza sociale, telefoni cellulari, proprietà, nomi e indirizzi attuali e precedenti ".

“Inoltre, l’azienda utilizza i dati sulle frodi provenienti dal suo “National Fraud Database”, un database di loro proprietà, che contiene nomi, indirizzi, recapiti telefonici di soggetti associati a frode (Christl, W., 2017).

Equifax, fondata ad Atlanta nel 1899, è una delle più antiche società di credit reporting, e fornisce “insights” e vende “prodotti e soluzioni su una serie di questioni alle piccole imprese e ai clienti aziendali e commerciali”. Ha sede in 24 paesi nel mondo, e “fornisce insights su 820 milioni di consumatori” (ACCC, 2023). I ricavi totali del 2023 sono stati 5,3 miliardi di dollari, di cui 2,3 miliardi nel segmento operativo dei “servizi di verifica” (e.g verifica del reddito, occupazione, istruzione, crimini commessi) degli individui negli Stati Uniti. Inoltre, offre “servizi informativi” per le imprese (e.g “soluzioni tecnologiche decisionali, servizi di gestione dell’identità, prevenzione di frodi, informazioni sui richiedenti mutui e servizi di marketing” sia negli Stati Uniti (1,7 miliardi di ricavi) che all’estero (1,2 miliardi di ricavi) (Equifax Annual Report, 2023). Equifax possiede dati su circa il 45 % degli investimenti in asset dei cittadini americani, che coprono riguardano dei potenziali “segmenti di targeting digitale” tra i quali ricchezza, auto, reddito, propensione alla spesa, prestiti, viaggi e tempo libero ecc. (Sherman, 2021).

Entrambe presentano similarità nella loro funzione principale di credit reporting, ma differiscono in termini di servizi e clientela alla quale si rivolgono. Experian, oltre al credit reporting, è fortemente attiva nell’analisi dei dati e del marketing con l’uso di tecnologie di machine learning per migliorare i risultati dell’analisi. Dunque, pone un forte accento sull’innovazione tecnologica e servizi di analisi avanzata. Oltre a imprese, istituzioni finanziarie e governi si rivolge anche ai consumatori individuali offrendo strumenti meno come il monitoraggio del credito e insight per il miglioramento del credit score. Viceversa, il business di Equifax è focalizzato sull’offerta di servizi per le imprese e meno sui servizi per i consumatori. Infatti, i clienti principali sono banche, imprese retailers e assicurazioni.

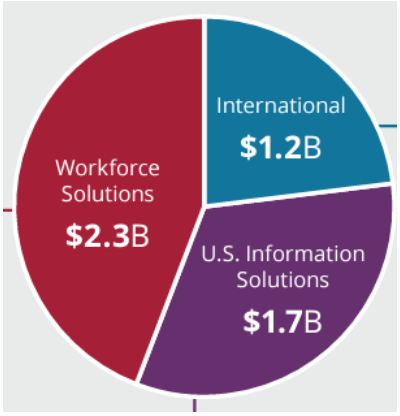

Figura 15: Equifax, ricavi consolidati 2023 per business unit.
Fonte:
(https://d1io3yog0oux5.cloudfront.net/_18cf1b1f549ee4879a576a94ae5052e1/equifax/db/2054/19426/annual_report/2023+Annual+Report.pdf)

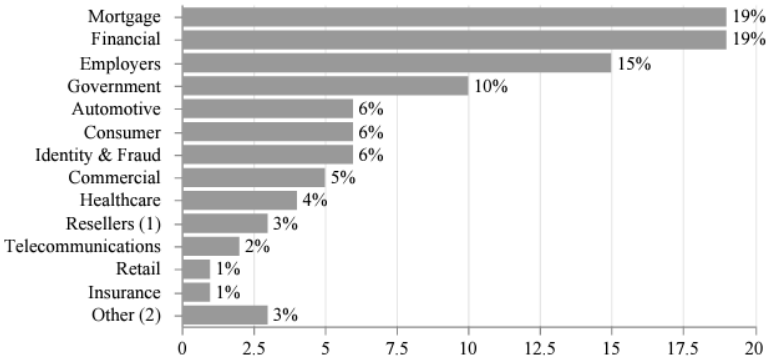

Figura 16: Equifax, percentuale dei ricavi 2023 per prodotti.
Fonte:
(https://d1io3yog0oux5.cloudfront.net/_18cf1b1f549ee4879a576a94ae5052e1/equifax/db/2054/19426/annual_report/2023+Annual+Report.pdf)

TransUnion insieme ad Equifax ed Experian è una tra le tre principali agenzie di **credit bureau** negli Stati Uniti. Come Equifax offre perlopiù servizi per le imprese (e.g gestione del rischio e analisi predittive sul marketing) e per gli istituti finanziari (e.g report sulla situazione finanziaria dei potenziali debitori). Minore è il focus sul consumatore, anche se il suo prodotto "IDVision", che incrocia i dati sull'identità personale e sulla reputazione (registri condivisi di frodi) con dati sull'identità digitale e sulle transazioni in tempo reale (e.g posizione, dispositivi mobili, utilizzo di app,

interazioni con siti web, impronte digitali del dispositivo, cronologia delle transazioni d'acquisto, indirizzi IP) (TransUnion Annual Report, 2023)

TransUnion possiede dati su 1 milione di consumatori in tutto il mondo collezionati da oltre 90.000 fonti (Christl, W., 2017). Nel 2023 i ricavi sono stati pari a 3,8 miliardi di dollari, di cui la metà circa derivanti dai servizi finanziari e di credit reporting e la rimanente parte da altri servizi quali marketing, retail, servizi per le compagnie assicurative ed healthcare (TransUnion Annual Report, 2023).

LexisNexis è una data company che lavora per le 50 migliori banche statunitensi e per più di 7500 autorità locali e federali. Offre prodotti e servizi quali "credit risk assesment, frodi, verifica dell'identità, customer information management" (LexisNexis website, 2024). La società è una partecipata di RELX PLC, una multinazionale britannica nell'ambito dell'analisi dati, precedentemente nota come Reed Elsevier.

Si stima che LexisNexis abbia dati sull'identità di 500 milioni di consumatori negli Stati Uniti. Il suo "TrueID system" ad esempio permette di verificare l'identità confrontandola con un database di 34 milioni di dati provenienti da 10.000 fonti. L'identità può dunque essere associata a: "carte di pagamento, assegni, carte fedeltà e altri dati del cliente". Si occupa anche di pubblicizzare dati relativi a crimini e indagini penali, avendo la capability di ricercare istantaneamente report dettagliati su individuo sfruttando i dati di 37 miliardi di fonti pubbliche. Tale compagnia ha la capacità di "filtrare e collegare miliardi di record per fornire un quadro più completo di un individuo", per trovare delle connessioni tra le persone e i loro beni, se parenti, soci d'affari oppure se il nome di un individuo è associato a una causa penale o una procedura di fallimento (Sherman, 2021). Inoltre, LexisNexis ha una "divisione governativa" che offre soluzioni per la prevenzione delle frodi in ambito dei servizi governativi e pubblici, come "i buoni pasto, assistenza sanitaria, pensione, prestiti agli studenti" (Christl, W., 2017).

Nel mercato del **marketing** e dell'**advertising** troviamo compagnie come Acxiom e Oracle. Acxiom è uno dei più famosi data broker presenti negli Stati Uniti "che colleziona, analizza e scambia un vasto ammontare di informazioni sui consumatori".

Fornisce data analytics per il marketing e in misura minore prodotti sulla mitigazione del rischio, verifica di identità e soluzioni per prevenire le frodi. Tra i clienti troviamo imprese nell'ambito dei "servizi finanziari, assicurazioni, telecomunicazioni, retail, healthcare, servizi governativi". Possiede **data elements** (vedi par. 2.2.2) su oltre 700 milioni di consumatori raccolti da migliaia di fonti, tra le quali altre compagnie e altri data broker. Acxiom opera principalmente negli Stati Uniti, ma ha anche delle sedi in Europa e nell'Asia-Pacifica. Dopo l'acquisizione della data company LiveRamp nel 2014, Acxiom facendo leva anche sulle capability della società acquisita, ha investito sforzi notevoli per integrare i suoi database con l'attuale universo del digital tracking e profiling. Infatti, ogni individuo presente nel database di Acxiom ha un identificativo unico che contiene nomi, indirizzi, recapiti telefonici, licenze di proprietà, veicoli posseduti, stato di salute, reddito, mutui, interessi politici, interessi religiosi, uso dei social media ecc., e che può essere incrociato con identificativi online e mobili (cookies and ID dei dispositivi) e offline (carte fedeltà o registri pubblici). In altre parole, Acxiom raccoglie in tempo reale dati come "pubblicazioni sui social media, visualizzazioni, click, ricerche sul web, iscrizioni, acquisti e rimborsi". Può addirittura capire "se qualcuno ha rimosso un item da un'acquisto online". Come detto in precedenza, Acxiom presenta numerosi fonti, tra le quali altri data broker come Equifax, Experian, TransUnion ed Epsilon. Acxiom può dunque combinare i propri dati sui consumatori con quelli forniti da terze parti (più di cento terze parti). Tali dati possono poi essere utilizzati dai clienti per categorizzare i consumatori e customizzare le offerte oppure per studiare il comportamento dei consumatori. Ad esempio, un consumatore può essere riconosciuto come "visitatore di un sito web" e gli si può proporre un'offerta personalizzata, basandosi solo sui dati del profilo a disposizione, senza necessità che l'utente faccia l'accesso al sito web. Acxiom inoltre ha promosso come suoi fornitori ufficiali compagnie come Samba TV e Crossix, e collabora con Google, Facebook e Twitter fornendo a loro servizi di tracciamento e profilazione dei consumatori sulla base dei dati raccolti da queste piattaforme.

Figura 17: Fonti, partners e servizi di Acxiom.
Fonte: (Christl, W., 2017)

Oracle, multinazionale del settore informatico fondata nel 1977 in California, vende prodotti come software per la gestione dei database, software per le aziende e sistemi di ingegneria cloud. Tuttavia, in questi anni è diventato uno dei maggiori broker di dati sui consumatori al mondo. Recentemente ha infatti acquisito diverse data companies come "Datalogix (in seguito ridenominata "Oracle Data Cloud"), che aggrega miliardi di dati sulle transazioni d'acquisto su oltre 50 catene alimentari e 1500 grandi rivenditori, AddThis che traccia in tempo reale 900 milioni di utenti su 15 milioni di siti web e 1 miliardo di utenti mobili, Crosswise che raccoglie dati sull'attività in miliardi di dispositivi e identifica quali PC, telefoni, tablet e TV vengono utilizzati da un singolo consumatore". Tali acquisizioni hanno permesse ad Oracle di ampliare la dimensione del suo business. Inoltre, Oracle "aggrega e analizza 700 milioni di messaggi sui social media al giorno provenienti piattaforme di social media, bacheche, blog, recensioni dei consumatori e piattaforme video". Considerando dunque le varie acquisizioni e i dati già in suo possesso, si stima che Oracle abbia oltre 30.000 attributi su oltre 2 miliardi di consumatori. I consumatori vengono divisi e ordinati in categorie sulla base di

caratteristiche come età, genere, occupazione, reddito, prestiti e ricerche online. Ad esempio, sfruttando le ricerche sul web su questioni come “aborto, legalizzazione delle droghe, matrimoni gay o strutture mediche”, riesce a tracciare e categorizzare gli individui. Si stima che riesca ad assegnare ai consumatori 50.000 categorie differenti. Oracle possiede un proprio data cloud in cui i clienti “possono caricare i loro dati sui consumatori, utenti web e utilizzatori di app, i quali si combinano con i dati di Oracle e dei suoi partner presenti nel cloud. Tali dati possono essere utilizzati dai clienti per fornire servizi di marketing e advertising, per tracciare le persone attraverso le piattaforme, per personalizzare le offerte e misurare il comportamento dei consumatori”. Oracle, sfruttando i caricamenti dati dei clienti sul cloud, riesce a tracciare e identificare in tempo reale le loro interazioni e i comportamenti attraverso la tecnologia “ID Graph”. Quando i clienti caricano i dati sul cloud, vengono identificati attraverso una “chiave di corrispondenza, che li identifica sia nello spazio online che offline”. In altre parole, tali chiavi sono degli pseudonimi, ovvero dei codici riferiti a nomi, indirizzi e-mail ecc, che una volta inviate dagli utenti a Oracle, vengono collegate, incrociate e mappate all’interno della rete di ID utente anonimi e statistici presenti in “GraphID”. Possiede inoltre un proprio spazio di mercato online chiamato “Oracle BlueKai” in cui le imprese possono vendere i propri dati sui consumatori. Oracle inoltre collabora con Facebook, fornendo dati, analytics tool e servizi per aiutarla a “categorizzare e ordinare”, sulla base dei dati raccolti da Facebook all’interno della piattaforma, gli utenti del social media. Oltre che dai suoi clienti, Oracle colleziona dati anche da altri data broker come: “Acxiom, Experian, TransUnion ed Equifax” e da fornitori di dati come “Vis-ualDNA, Lotame” e da grosse companie come “Visa e Mastercard” (Christl, W., 2017).

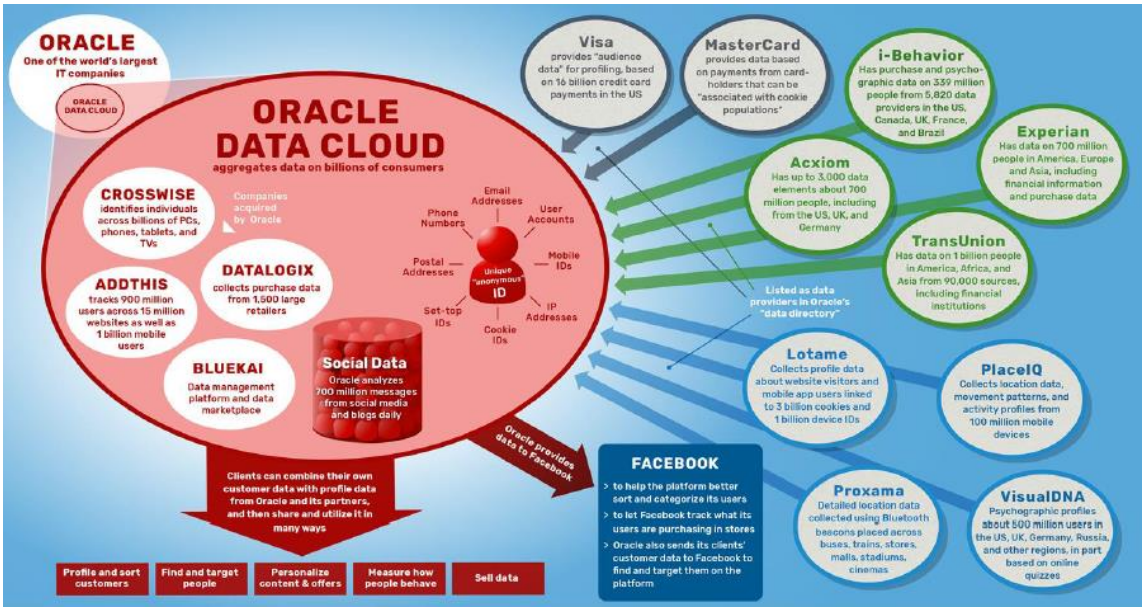

Figura 18: Ecosistema di Oracle.
Fonte: (Christl, W., 2017)

Tra altri data broker presenti nel mercato ricordiamo:

- Corelogic: una società di analisi che tratta principalmente dati sulle proprietà immobiliari per costruire report sull'accessibilità all'acquisto delle abitazioni e report sull'andamento del settore delle costruzioni e anche dati sui risultati delle vendite e delle aste e modelli di valutazione (ACCC, 2023). Possiede dati su oltre 1 miliardo di registrazioni immobiliari negli Stati Uniti, e anche liste di proprietà immobiliari, registri fiscali e dati sulla valutazione delle case, coprendo il 99% delle proprietà degli Stati Uniti (FTC, 2014). Tra i principali clienti vi sono aziende che operano nel settore immobiliare, nel settore bancario e finanziario, nell'edilizia e nel settore pubblico.
- Ilion: fornisce prodotti e servizi di analisi e dati, tra cui soluzioni per la gestione del rischio e analisi per il marketing (ACCC, 2023). Fa parte di Dun & Bradstreet Worldwide Network, un'alleanza di fornitori di informazioni sul business che opera in 220 paesi. Sfruttando questa rete, Ilion riesce a combinare le proprie conoscenze con i dati di oltre di 500 milioni di aziende nel mondo.
- Verisk: è una società di data analytics che opera nel settore delle assicurazioni, commerciale e dei servizi finanziari. Utilizza i dati provenienti da transazioni commerciali, registri pubblici e liste di proprietà per aiutare i propri i clienti a

valutare meglio il rischio e migliorare le prestazioni. Ad esempio, nel settore delle assicurazioni fornisce modelli predittivi e tool per valutazione del rischio nelle polizze auto e assicurazioni sanitarie, utilizzando il suo "Verisk Data Exchange", un database contenente dati sulle auto, informazioni personali sui proprietari e sui sinistri commessi. Tra i suoi prodotti offre anche una funzione di "Reverse Phone Append", che permette di ottenere dati su un individuo semplicemente inserendo il suo numero di telefono (Sherman, 2021). I suoi ricavi nel 2023 sono stati pari a circa 2,6 miliardi di dollari di cui la quasi totalità nei servizi per le assicurazioni, mentre nel 2022 dei 2,4 miliardi di dollari di ricavo vi era una quota parte attribuita a servizi per il settore dell'energia e servizi finanziari, pari a rispettivamente 22 e 37 milioni di dollari (Verisk Annual Report, 2023)

- Nielsen: è una compagnia statunitense operante in 55 paesi nel mondo e si occupa di misurazione dell'audience, insight sul consumo del media e ottimizzazione del marketing. È famosa per il Nielsen Ratings, un indice di misurazione dell'audience televisivo. Raccoglie dati da varie fonti come smart TV e dispositivi, dai social media, da panel di consumatori che accettano di sottoporsi a monitoraggio delle abitudini mediatiche e da altre compagnie. I suoi ricavi nel 2021 sono stati pari a 3 miliardi e mezzo di dollari (ACCC, 2023).

2.2.1 Le fonti utilizzate per estrarre i dati

I data broker possono acquisire i dati da diverse fonti, delle quali spesso non rivelano i dettagli. Le fonti possono essere suddivise in:

- **fonti pubbliche**;
- **fonti private** (dati non pubblicamente disponibili, ma ottenuti con uno scambio privato);

- **tracciamento online**;

I **dati pubblici** sono raccolti generalmente attraverso tecniche di "**web crawling**[8]" e "**web scraping**[9]"(software progettati per raccogliere automaticamente i dati da Internet) oppure acquistati da altri data broker "specializzati nella digitalizzazione di particolari tipi di documenti". Tra le fonti pubbliche figurano (Rieke, A. et al., 2016):

- **registri pubblici**: come "report di proprietà" e quindi "atti notarili, atti di acquisto o passaggio di proprietà, visure catastali, informazioni identificative sul proprietario e informazioni sull'immobile", "registri del tribunale" che contengono "casellario giudiziale, azioni e procedimenti civili a carico dell'individuo, condanne penali, documenti di nascita, matrimonio, divorzio e certificati di morte". Altre fonti pubbliche sono: "licenze professionali, licenze ricreative (caccia e pesca ad esempio), liste elettorali, registri dei veicoli a motore, patenti di guida". Un ruolo importante è giocato dagli enti pubblici che possono essere una fonte considerevole di dati. Ad esempio, lo U.S. Census Bureau può fornire informazioni demografiche come "etnia, età, livello di istruzione, composizione del nucleo familiare, reddito" su particolari zone della città oppure la Social Security Administration può fornire informazioni sui "nomi dei consumatori, social security number e data di morte" (FTC, 2014);
- **fonti commerciali**: ad esempio annunci pubblicitari su siti web, riviste, programmi di fidelizzazione dei clienti oppure elenchi telefonici pubblici. Tra

[8]Il **web crawling**, a differenza del **web scraping** che si focalizza sull'estrazione di dati specifici da una pagina web, ha un raggio d'azione più ampio e mira a esplorare il web per poter scoprire e indicizzare il maggior numero possibile di pagine, contenuti della pagina, metadati e URL. Viene avviato un apposito script, (detto "crawler"), che naviga automaticamente tra i link ipertestuali presenti in una pagina, raccogliendo e indicizzando l'intero contenuto della pagina. È utilizzato su larga scala e permette di raccogliere dati, strutturati e no, da milioni di pagine web in modo automatizzato. Questo processo automatizza la scoperta e l'analisi di un grande numero di pagine web, in tempi rapidi, e permette di accedere a un grande numero di informazioni pubblicamente disponibili su Internet (Reviglio, U., 2022).

[9]Il **web scraping** è una tecnica informatica che utilizza dei programmi software che simulano la navigazione sul Worl Wide Web. Viene avviato uno script che invia una richiesta HTTP ad un sito, scaricano il codice sorgente della pagina e poi scansionano ("scrapano") il contenuto per estrarre le informazioni desiderate. L'obiettivo è raccogliere dati specifici e mirati da una pagina web in modo automatico. Questo approccio permette di aggregare e raccogliere, in poco tempo, tutti i dati desiderati. Generalmente i dati, al momento della raccolta automatica, si presentano in modo non strutturato (notizie sul web, immagini, video ecc.), quindi andranno poi trasformati in dati strutturati(tabelle) per la memorizzazione in un database (Reviglio, U., 2022).

le fonti commerciali ci sono anche i “retailers” e “le aziende di cataloghi” da cui è possibile ottenere delle informazioni sulle transazioni avvenute per gli acquisti, ad esempio il tipo di prodotto acquistato (elettrodomestici, vestiti, articoli sportivi ecc.), l’importo pagato, la data dell’acquisto e la modalità di pagamento. Le aziende di cataloghi e i retailers collezionano informazioni su nome dei clienti, indirizzi e storico degli ordini che diventano delle fonti potenziali per i data broker (Rieke, A. et al., 2016);

- **media e social network**: si possono reperire informazioni dai profili Linkedin, Facebook, You Tube, dalle discussioni sui blog ecc.

I data broker possono ottenere dati da **altre aziende o entità private** o da altri **data broker** attraverso uno scambio regolamentato da un contratto scritto (Rieke, A. et al. 2016). I contratti di scambio con le fonti possono assumere diverse sfumature, si può ad esempio acquisire la proprietà dei dati con un classico contratto di compravendita oppure è possibile ottenere la licenza all’utilizzo di questi dati per un determinato periodo. Tali contratti contengono generalmente “una descrizione dei dati forniti al data broker, il metodo di trasferimento dei dati, la frequenza degli aggiornamenti e le eventuali restrizioni all'uso dei dati” (FTC, 2014). In alcuni contratti possono essere presenti delle clausole contrattuali che attestano che i dati “sono stati ottenuti rispettando le normative vigenti” (Rieke, A. et al. 2016). Le principali fonti private sono:

- **rivenditori al dettaglio**, i quali“vendono informazioni sugli acquisti dei loro clienti, la frequenza e la modalità di acquisto”. Ad esempio, Datalogix, data company acquisita da Oracle, possiede dati su oltre “1 trilione di spesa dei consumatori su 1400 marchi leader” (Rieke, A. et al., 2016);
- **istituzioni finanziarie** come “banche, istituti di credito e assicurazioni” che rilasciano informazioni sui loro clienti ai data broker in cambio di un accesso a documenti come credit report e credit score.
- **proprietari dei siti web, applicazioni** e **piattaforme digitali**: ci sono centinaia di migliaia di siti web che condividono dati con altre compagnie, inclusi i data broker. Tali informazioni sono generalmente liste di persone che si sono registrate al sito.

Inoltre, i data broker ottengono informazioni anche dalle piattaforme digitali (siti di vendita al dettaglio o testate giornalistiche). Infatti, ogni qual volta si effettua la registrazione o il login ad un sito web o ad una piattaforma digitale, vengono rilasciate un gran numero di informazioni come nome, indirizzo e-mail, numero di telefono, età ecc (Rieke, A. et al., 2016). Va evidenziato che, dietro alcune piattaforme, si nascondono in realtà dei data broker.

A volte i dati sono forniti dagli individui stessi che, quando utilizzano un'app, devono fornire l'accesso una serie di dati (GPS, audio, fotocamera, contatti personali) per poter continuare ad utilizzare il servizio. E i proprietari di applicazioni spesso condividono questi dati con altre entità, inclusi i data broker (Reviglio, U., 2022);

- **sviluppatori di app**, attraverso l'uso di "**SDKs**" (Software Development Kits), ovvero dei tools che i data broker forniscono gratis agli sviluppatori di app. Tali tools aiutano lo sviluppatore ad accelerare il processo di implementazione della app. In cambio i data broker richiedono di poter accedere ai dati sugli utilizzatori dell'app (ACCC, 2023).
- altri data broker (Rieke, A. et al. 2016).

Infine, il terzo modo in cui ottenere informazioni è quello di **tracciare** il comportamento online di un utente sul web. Ciò può avvenire in diversi modi, ad esempio attraverso il tracciamento dei **cookies**[10] (ACCC, 2023) o il **browser fingerprinting**[11]. I cookies sono

[10] Si è soliti suddividere i cookies in **cookies di prima parte** e **cookies di terza parte**. I primi sono trasmessi dal proprietario del sito web, mentre i secondi da terzi soggetti diversi dal proprietario del sito, il quale ne ignora generalmente la quantità raccolta e le finalità per le quali saranno poi sfruttati. I cookie di terze parti consentono alle aziende di fare una profilazione "cross site" (attraverso molteplici siti) e costruire dunque un profilo più dettagliato dell'utente. Sono fondamentali nella "Pubblicità Programmatica" e nel "Retargeting" (vedi par.2.2.3). Fonte: https://www.digital4.biz/marketing/advertising/cookie-di-terza-parte-e-prima-parte-cosa-sono-come-cambia-la-pubblicita-online/

[11] È un metodo che permette l'identificazione dell'utente o del dispositivo anche quando i cookie sono disattivati. Si basa sul fatto che, quando effettuiamo la connessione ad un sito web, tale sito web utilizza degli script nascosti e raccoglie informazioni sulla versione del browser, sul sistema operativo, indirizzo IP, plug-in attivi, la lingua utilizzata, il fuso orario, informazioni sullo schermo ecc. Queste informazioni, in genere condivise di default per far funzionare app e siti web correttamente, "permettono di costruire un profilo personale dell'utente e sono quindi utilizzate per identificare gli utenti, proprio come farebbe un'impronta digitale". "Meno comuni sono le impostazioni e i dati raccolti, maggiori sono le probabilità di essere riconoscibili e rintracciabili". Fonte: "https://www.agendadigitale.eu/sicurezza/privacy/controllati-online-col-browser-fingerprinting-cose-e-come-proteggerci/#Cose_e_come_funziona_il_browser_fingerprinting"

dei piccoli file di testo che, quando un utente visita un sito web, vengono archiviati nella memoria del browser dell'utente a opera del server del sito web. Tali file possono contenere informazioni sul comportamento di navigazione dell'utente come siti visitati, tempo trascorso su un sito ecc. I cookies vengono memorizzati nel browser, permettendo così al sito web di ricordare le informazioni e le preferenze dei visitatori, che navigheranno nuovamente sul sito. Ad esempio, è grazie ai cookie che vengono salvate le preferenze di lingua o di visualizzazione di un sito web e che un carrello salvi i prodotti aggiunti anche tra una sessione e l'altra. Generalmente vengono cestinati una volta che è finita la sessione web, anche se alcuni tipi di cookies rimangono sul browser fino ad una data di scadenza. Da un lato consentono un miglioramento della navigazione sul sito, conservando ad esempio il login o la preferenza di una lingua, dall'altro possono essere utilizzati per "monitorare" le attività online dell'utente, come ad esempio i click sulle inserzioni pubblicitarie, col fine di creare un profilo dettagliato delle abitudini e delle preferenze del visitatore a scopi di marketing (Polimeni, A., 2022). Le origini del posizionamento sul web dei cookies sono da attribuire agli online advertisers, Ad esempio, DoubleClick, la piattaforma di advertising di Google acquistata nel 2007, permetteva agli inserzionisti sia di acquistare spazi pubblicitari sulla piattaforma che di acquistare il diritto di fare pubblicità mirata ad un determinato tipo di utente (pratica nota come "pubblicità programmatica") (Reviglio, U., 2022).

Con la normativa vigente, le informazioni provenienti dai cookies di tracciamento possono essere trattati solo con l'espresso consenso dell'utente. Tuttavia, si stima che "meno del 30% degli utenti cancelli i cookies" contribuendo a quello che è definito il "paradosso della privacy". Infatti, gli utenti, pur riconoscendo l'importanza della privacy, "raramente agiscono per proteggerla", nonostante gli strumenti messi a loro disposizione dai legislatori, come la Cookie Law e il GDPR (Reviglio, U.,2022).

Tipicamente i data broker verificano l'affidabilità e l'accuratezza dei dati forniti dalle loro fonti. In alcuni casi, come per fonti commerciali, si basano sulla reputazione dell'azienda nel settore. In altri casi, tuttavia, vengono fatte delle vere e proprie verifiche sulla legittimità, e dunque l'ispezione del "sito web della fonte, i termini di utilizzo, i

metodi di raccolta dei dati, l'informativa sulla privacy, le pratiche sulla privacy e la conformità normativa" e sull'accuratezza dei dati. In questo caso, prima dell'acquisizione, i dati vengono collocati in un'"area di attesa" per certificare che tali dati siano "internamente coerenti, corroborati da altre fonti e includono una porzione sufficientemente ampia della popolazione". Inoltre, ci può essere anche una comparazione con altre fonti ritenute di qualità dai data broker oppure i dati possono essere dati in pancia a sistemi automatici che verificano le "deviazioni materiali nei dati e le fonti che le causano". Così in caso di discordanza tra le fonti, si tenderà a preferire coloro che hanno dimostrato di essere più affidabili nel tempo (FTC, 2014).

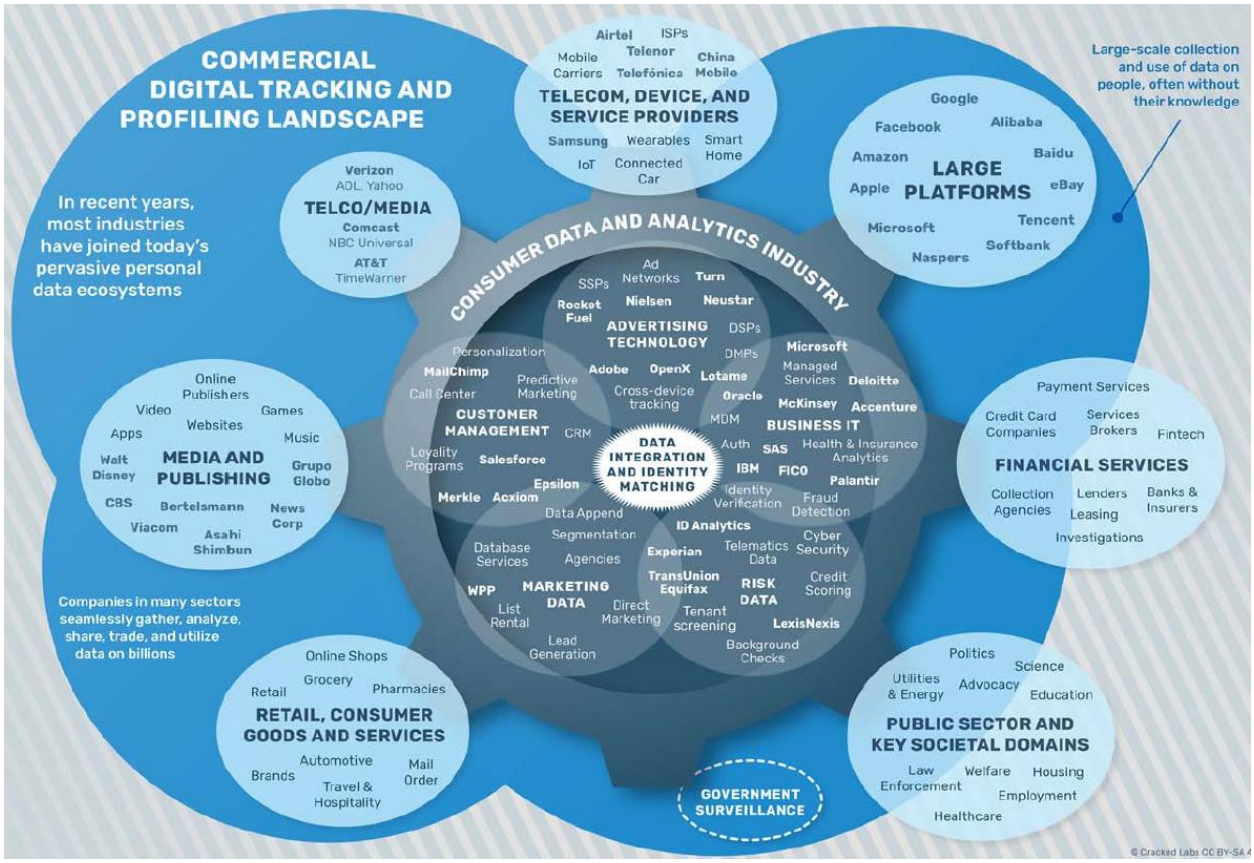

Figura 19: Alcune fonti dei data broker.
Fonte: (Corporate Surveillance in Everyday Life, 2017)

2.2.2 Le tipologie di prodotti offerti

I data broker vendono diversi tipi di prodotti e servizi sempre più adattati alle specificità dei loro clienti, i quali attribuiscono a quest'ultimi un ruolo centrale nei problemi decisionali delle loro aziende (Reviglio, U., 2022). Nel mercato dei data broker è possibile distinguere tre categorie di prodotti:

- prodotti per la **mitigazione del rischio**;
- prodotti per il **marketing** e l'**advertising**;
- **people search product**;

È bene, anzitutto, premettere che vi è differenza tra **data elements** o **raw data**, ovvero "dati grezzi" che vengono raccolti a partire da diverse fonti pubbliche e private e che non subiscono elaborazioni e, **derived data elements**, i dati creati a partire da deduzioni ed elaborazioni dei dati grezzi. Degli esempi di data elements sono: nome, indirizzo, età, etnia, reddito. Un data broker, partendo ad esempio da un consumatore che ha una patente nautica, può dedurre che tale consumatore ha un interesse verso il mondo nautico oppure che un consumatore ha interesse per l'innovazione tecnologica in quanto quest'ultimo ha sottoscritto un abbonamento alla rivista Wired. Questi sono tutti esempi di derived data elements (FTC, 2014).
Nel mercato della **mitigazione del rischio** si trovano prodotti come **report di credito**[12] che vengono acquistati da banche e altri tipi di intermediari finanziari per aumentare l'accuratezza del credit score e valutare il merito creditizio del richiedente prestito. In questo modo le banche riescono ad effettuare con maggiore precisione lo screening dei potenziali clienti (Rieke, A. et al., 2016). Oppure tali report possono essere acquistati da società che erogano prestiti per auto o mutui, fornitori di servizi di telefonia mobile e agenzie di recupero crediti. Il mercato del rischio negli Stati Uniti è molto vasto e si

[12] I report di credito sono documenti che contengono informazioni sulla storia creditizia del consumatore, storico dei debiti saldati, dettagli sui conti correnti come prestiti, mutui e carte di credito, fallimenti o procedimenti giudiziari a carico dell'individuo. Le agenzie di credit reporting le utilizzano per prevedere il merito creditizio con una qualità superiore rispetto a quella che si avrebbe con l'uso di dati tradizionali (Christl, W., 2017).

trovano anche **report sugli affittuari** che contengono informazioni sulla cronologia dei pagamenti delle rate dell'affitto, sugli sfratti e sui precedenti penali dell'individuo che vengono utilizzati da operatori nel settore dell'immobiliare o proprietari di immobili da affittare, **report sull'occupazione** contenenti dettagli sull'impiego, il salario, lo storico dei lavori precedenti, istruzione, licenze professionali, risultati dei test antidroga e alcool e così via. Sono generalmente adoperati da datori di lavoro, agenzie di somministrazione di lavoro, società di consulenza e agenzie governative. In questo mercato si trovano imprese come Experian, Equifax e TransUnion e LexisNexis. (Christl, W., 2017)
A sua volta nel mercato della mitigazione del rischio, i prodotti esistenti si suddividono in due categorie: (FTC, 2014)

- **verifica dell'identità**: ovvero prodotti che aiutano le imprese clienti a confermare l'identità di una persona con cui effettueranno una "transazione". Tali prodotti presentano configurazioni diverse. Alcuni assegnano un "risk score" calcolato sulla base della rischiosità dell'individuo con cui si interfacciano. Più alto è il punteggio, maggiore è il rischio associato. In genere per creare questi prodotti viene utilizzato l'SSN[13] degli individui. Un punteggio alto fa scattare dei "codici esplicativi" che potrebbero rilevare che l'SSN è associato ad esempio "ad una persona deceduta, a casi di frode oppure l'indirizzo collegato all'SSN è diverso da quello indicato dal cliente. Altri prodotti forniscono "un'ulteriore livello di autenticazione" sotto forma di un quiz, la cui risposta è conosciuta solo dal consumatore. Ciò permette di individuare facilmente un "ladro d'identità". Se il "risk score" è alto, "il cliente del data broker può richiedere al consumatore di rispondere correttamente a cinque domande su sei, nel caso di punteggio basso può essere richiesto un tre su cinque". Inoltre, i data broker, offrono anche prodotti con verifica "match/no match", ossia viene notificata una conferma se le

[13] "Social Security Number". È un codice in possesso di ogni cittadino americano che serve per scopi fiscali e per accedere ad alcuni servizi pubblici. È simile al codice fiscale in Italia. Fonte: https://it.wikipedia.org/wiki/Social_security_number

informazioni fornite dal cliente sono in linea con quelle presenti nei loro archivi. (FTC, 2014)

- **prevenzione delle frodi**: strettamente collegati con la verifica dell'identità, tali prodotti aiutano i clienti dei data broker a identificare delle potenziali frodi. Ad esempio, ci sono prodotti che permettono di verificare se un indirizzo e-mail esiste veramente, da quanto tempo è stato creato e se ha "se ha una storia di transazioni ad esso collegate", oppure i data broker possono tenere traccia degli indirizzi dei consumatori per individuare degli schemi tipicamente associati a tentativi di truffa ("l'indirizzo di consegna non è associato al consumatore indicato o un SSN è associato a diversi indirizzi") (FTC, 2014). Nel mercato del **marketing** invece, i prodotti possono essere suddivisi in tre categorie: (FTC, 2014)
    - **marketing diretto**
    - **marketing online**
    - **marketing analytics**

È possibile suddividere a sua volta il **marketing diretto** in due categorie:

- **"data append"**: questo tipo di prodotti "aiutano le aziende a conoscere meglio i propri clienti". L'azienda fornisce al data broker alcune informazioni sui propri clienti come nome e indirizzo e-mail, e in base alle richieste dell'azienda, il data broker integra questo set di informazioni con quelle a propria disposizione. Ad esempio, l'azienda può fornire il "nome e l'indirizzo del cliente e il broker di dati può aggiungere il numero di telefono fisso o l'indirizzo e-mail" o ancora l'azienda fornisce "il numero di telefono cellulare del cliente il data broker aggiunge il nome e l'indirizzo del cliente". I dati aggiuntivi possono includere anche dati demografici, abitudini d'acquisto e interessi del cliente. Questo genere di prodotti serve alle aziende "per arricchire i dati a propria disposizione sui clienti e quindi conoscerli meglio". Possono essere utilizzati dalle aziende per campagne di marketing diretto come "telemarketing ed e-mail marketing" ma anche per "segmentazione dei clienti

e marketing mirato". Tra le informazioni che i data broker possono "aggiungere" ci sono anche una serie di **derived data elements** come: "interesse per la tecnologia, reddito familiare, occupazione, presenza di figli, abitudini di investimento, presenza di fumatori in famiglia, utilizzo della carta di credito, bambini in età scolare, neo-genitori, affiliazione politica e religiosa" (FTC, 2014).

- **marketing lists**: servono a identificare i consumatori che condividono particolari caratteristiche o attributi come ad esempio: "persone che vivono con almeno due figli, persone che sono donne e possiedono una determinata marca di auto, persone interessate al diabete, famiglie con fumatori, persone che vivono in determinate aree, persone con una determinata fascia di reddito". In genere l'azienda "identifica gli attributi che desidera trovare in un pubblico di consumatori e l'intermediario di dati fornisce un elenco di consumatori con tali attributi". Ad esempio, un'azienda che vende prodotti per l'infanzia potrebbe chiedere al data broker una lista di neogenitori o un'azienda che vende articoli sportivi potrebbe richiedere una lista di persone interessate al fitness. Tali liste possono contenere nomi, numeri di telefono e indirizzi e-mail dei clienti che servono per le campagne telefoniche di marketing e per l'e-mail marketing. Se invece le aziende desiderano dei dati "più consistenti" per personalizzare al meglio le loro campagne di marketing, nelle marketing lists possono essere inseriti dei derived data element ("età, reddito familiare, patrimonio netto, occupazione ecc.). Quindi ad esempio, un fornitore di prodotti gourmet può chiedere al data broker una lista di consumatori interessati all'alta cucina in una particolare regione, con un determinato reddito familiare e una determinata fascia di età. (FTC, 2014)

Il settore dell'**online advertising** negli ultimi anni ha avuto un grande sviluppo. Infatti, nonostante il complesso ecosistema, fatto da "aziende e tecnologie diverse che interagiscono fra di loro", solo una piccola parte sia "sia in termini di complessità che

di volume" è perfettamente visibile all'utente (Christl, W., 2017). In questo mercato si possono individuare tre diverse categorie di prodotti (FTC, 2014):

- "**targeting di registrazione**": i cosiddetti "siti di registrazione", ovvero in cui è richiesta la registrazione per ottenere il servizio, come siti di vendita al dettaglio, informazione giornalistica, viaggi, social media, possono chiedere una mano ai data broker per "promuovere i prodotti attraverso un'esperienza utente più personalizzata". Ad esempio, un sito di viaggi può inviare al data broker una lista degli utenti registrati e il data broker può fornire gli interessi e le preferenze di viaggio di quegli utenti specifici. Grazie a queste informazioni, il sito di viaggi può offrire dei pacchetti personalizzati ai suoi utenti. Oppure la lista degli utenti registrati può essere utilizzata per vendere spazi pubblicitari sul sito dell'agenzia di viaggi. Ad esempio, il data broker, una volta ricevuta la lista, informa l'agenzia di viaggi che la maggioranza dei suoi utenti ha interessi per le moto e per prodotti per l'igiene della casa. Allora l'agenzia di viaggi può vendere spazi pubblicitari sul proprio sito a rivenditori di motocicli e di prodotti per l'igiene della casa (FTC, 2014).
- **targeting collaborativo:** mentre nel **targeting di registrazione** il data broker ha come unico cliente il sito di registrazione, in questo caso il data broker offre il proprio servizio a due clienti, il "sito web di registrazione" e "l'inserzionista che è alla ricerca di uno spazio pubblicitario su un sito di registrazione". "Il sito web di registrazione fornisce al data broker un elenco dei suoi utenti, mentre l'inserzionista fornisce al data broker il suo elenco di clienti e potenziali clienti". Il data broker, dopo aver analizzato i dati, può comunicare all'inserzionista la convenienza o meno di acquistare uno spazio pubblicitario sul sito web di registrazione. Ad esempio, un'inserzionista specializzato nella vendita al dettaglio di abbigliamento sportivo può richiedere uno spazio pubblicitario (tipicamente un banner) su un sito di viaggi ma inizialmente non sa se la sua clientela target visita abitualmente il sito. Allora l'inserzionista manda al data broker una lista dei suoi clienti e potenziali clienti, contenente nome e indirizzo

e-mail. Il sito di viaggi manda al data broker la lista degli utenti registrati, contenente sempre nome e indirizzo e-mail. Il data broker, incrociando i dati dell'inserzionista e del sito di viaggi, potrebbe identificare quali clienti sono registrati nel sito di viaggi. Qualora il numero di clienti registrati nel sito fosse ritenuto sufficiente dall'inserzionista, quest'ultimo fornisce al data broker l'inserzione che vorrebbe mostrare che a sua volta viene girata al sito di viaggi. A questo punto il sito mostra l'inserzione (FTC, 2014).

- **onboarding:** i data broker possono combinare i **dati offline** forniti dai clienti come nomi, storico degli acquisti, programmi fedeltà, con **dati online**, ovvero provenienti da ricerche sul web, e-commerce e interazioni sui social media. I dati offline attraverso identificatori univoci (e-mail o numero di telefono) vengono associati a profili digitali come i cookies, i quali memorizzano il comportamento di navigazione dell'utente come pagine visitate e tempo di navigazione su una determinata pagina. A questo punto, aggiungendo i dati offline, è possibile creare un profilo digitale completo, che consente agli inserzionisti di "utilizzare le attività offline dei consumatori per determinare quali pubblicità proporre loro su Internet" e di fatto "rivolgersi ai consumatori ovunque sul web". In poche parole, l'onboarding combina dati online e offline per creare dei profili digitali più dettagliati, per rendere più accurato il targeting pubblicitario e aiutare dunque gli inserzionisti.

Di solito, le aziende forniscono dei dati sui loro clienti ai data broker, i quali utilizzano questi dati per "identificare un pubblico di consumatori che condividono particolari caratteristiche e trovare tali consumatori su Internet per la distribuzione di annunci pubblicitari mirati". L'onboarding prevede, di solito, tre fasi: **segmentazione**, **matching** e **targeting**.

Il primo step inizia col cliente che chiede al data broker di trovare dei consumatori con particolari caratteristiche. Il data broker può già avere delle segmentazioni nel suo database o può crearle al momento in base alla richiesta del cliente. Ad esempio, un rivenditore al dettaglio di abbigliamento vuole lanciare una nuova linea basata sugli abiti

di lusso. Allora il rivenditore può dare la propria lista dei clienti al data broker, che la confronterà con i segmenti in suo possesso, come “acquirenti sofisticati” o “persone interessate all'abbigliamento di fascia alta”, per prevedere quali clienti del retailer potrebbero essere interessati alla nuova linea di abbigliamento e aiutare così il retailer a lanciare una campagna di marketing mirata. Oppure se il data broker non ha alcun segmento nel suo database, può crearne uno nuovo basandosi sui vincoli stabiliti dal rivenditore al dettaglio, ad esempio “donne residenti nel codice postale 12345”. È possibile anche che il data broker dia al retailer l’accesso ai propri segmenti, per permettergli di trovare nuovi clienti.

Lo step seguente è il **matching**, in cui il data broker compra dai siti web le liste di utenti registrati e le confronta con i consumatori identificati attraverso il processo di segmentazione. Quando trova una corrispondenza, il data broker può arricchire il profilo del consumatore con le informazioni offline e online. Infine, abbiamo la fase di **targeting**, in cui il data broker inserisce un cookie nel browser dei consumatori di cui ha trovato una corrispondenza, come descritto sopra. Il cookie include “le informazioni che il data broker ha aggiunto al profilo del consumatore” di cui ha trovato corrispondenza. Il cookie viene inviato nel momento in cui il consumatore si connette al sito, viene infatti notificata, dal sito al data broker, l’avvenuta connessione dell’utente. A questo punto, il data broker può fare pubblicità al consumatore per tutto il tempo che il cookie starà sul suo browser. Dunque, può agire come una rete pubblicitaria vera e propria acquistando spazi pubblicitari sui siti web oppure può acquistare tali spazi pubblicitari e cederli alle aziende interessate (FTC, 2014).

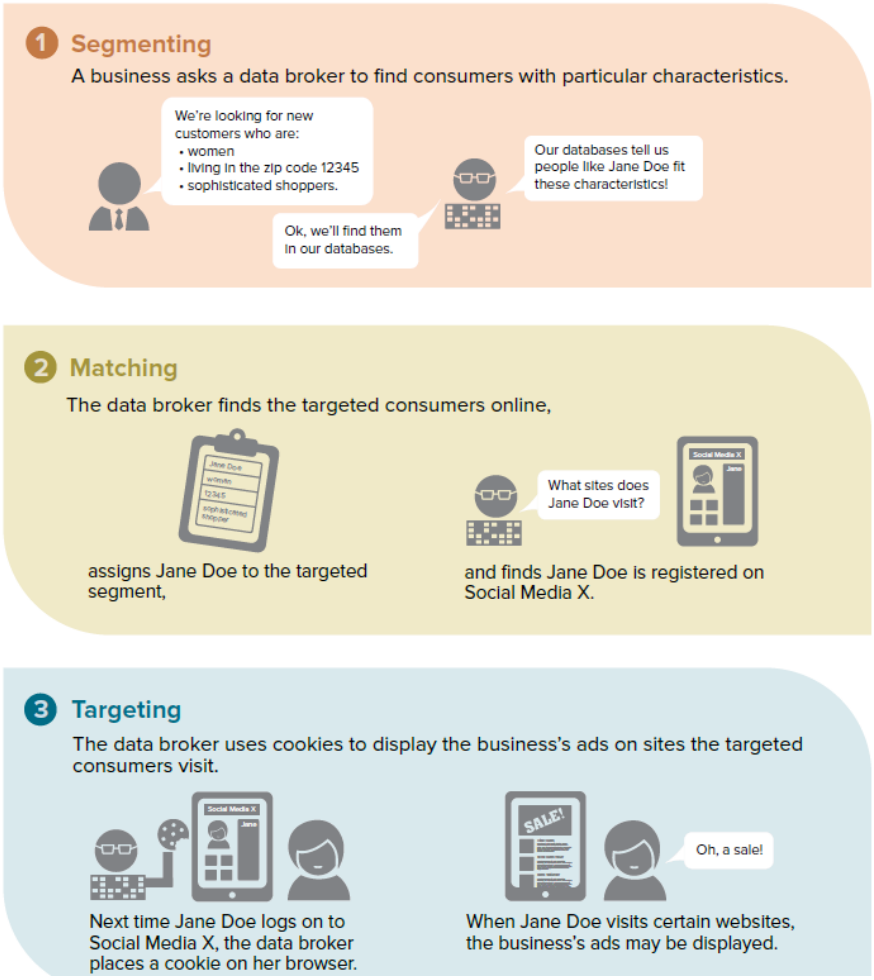

Figura 20: Processo di onboarding.
Fonte: (FTC, 2014)

Il processo di onboarding può inoltre aiutare le campagne di **pubblicità programmatica** e di **retargeting** (per la definizione vedi par. 1.3.2). Il primo è un'asta automatizzata, tra inserzionisti ed editori (proprietari dei siti web in cui saranno mostrati gli annunci pubblicitari), per l'acquisto di spazi pubblicitari online. Generalmente quando un utente visita una pagina web invia agli inserzionisti dei metadati sul contenuto della pagina, sul suo profilo e anche un'identificativo (cookie di terza parte, indirizzo IP, identificatori di sessione, account utilizzato per il login ai social media) che consente di tracciare. Così gli inserzionisti interessati a mostrare un annuncio a questo particolare utente fanno un'offerta. Le offerte vengono fatte su una piattaforma digitale, detta Ad Exchange, in cui inserzionisti ed editori scambiano gli spazi pubblicitari. Tale piattaforma funge da

intermediario tra il lato della domanda, detto “DSP” (Demand-side platforms) e rappresentato dagli inserzionisti, e il lato dell’offerta detto “SSP” (Supply-side platforms. L’inserzionista che fa l’offerta più alta vince l’asta e ottiene lo spazio pubblicitario. È bene sottolineare che le offerte sono in “tempo reale” e avvengono nel giro di pochi millisecondi dal momento in cui inizia il caricamento del sito web. Si tratta di un processo molto efficiente in quanto gli annunci vengono mostrati all'utente giusto e in “tempo reale”. Col **retargeting** invece, l’impresa può rivolgersi a una lista di persone che sono già clienti suoi clienti e utilizzare le informazioni già in suo possesso per indirizzare a essi offerte specifiche su Internet. Ad esempio, un istituto di credito potrebbe rivolgersi ai clienti in difficoltà finanziaria per indirizzare, attraverso Internet, specifici annunci riguardanti una carta di credito subprime (con tassi di interesse alti) oppure un hotel può utilizzare la lista dei suoi clienti più fedeli per invitarli a soggiornare in hotel (magari con uno sconto), sapendo che tali clienti hanno un alto potenziale di spesa (FTC, 2014).

Il settore dell’online advertising è dominato da Google, grazie anche alla propria piattaforma di advertising DoubleClick, che fa parte del sistema Google Marketing Platform. Tale piattaforma è integrata con gli strumenti di Google e consente lo scambio di spazi pubblicitari tra inserzionisti ed editori. Ciò ha rafforzato la centralità di Google nel settore e lo ha reso un intermediario importante per questo genere di transazioni (Christl, W., 2017). Secondo uno studio del 2021, Google nel 2019 rappresentava il 31% della pubblicità digitale negli Stati Uniti, seguita da Facebook col 23% con entrate da advertising pari a rispettivamente 39 e 31 miliardi di dollari (Calderini, B., 2021).

Infine, abbiamo il settore dei **marketing analytics**, ovvero dei prodotti che mirano a prevedere e ottenere approfondimenti sul comportamento dei consumatori, a fare inferenze sulle loro abitudini e a perfezionare le campagne pubblicitarie. Questi prodotti vengono realizzati partendo da una “**profilazione**” dei consumatori, svolta con l’ausilio di tool automatizzati che usano i dati personali per analizzare e prevedere aspetti relativi ad una persona fisica, tra i quali “il rendimento professionale, la situazione economica, la salute, le preferenze personali, gli interessi, l'affidabilità, il comportamento,

l'ubicazione o gli spostamenti" (art. 4 par. 4 GDPR). I data broker, dunque, attraverso la "profilazione" dei consumatori creano i prodotti "marketing analytics" che, possono aiutare i loro clienti a trovare il canale ottimale attraverso il quale lanciare la campagna pubblicitaria (ad esempio social media, giornali, televisione, e-mail ecc.) oppure la zona geografica in cui concentrare la campagna. Questi prodotti permettono inoltre, agli acquirenti, di valutare se una campagna pubblicitaria su una determinata piattaforma avrà gli effetti desiderati o di stimare l'impatto di una campagna dopo il suo lancio.

Un'altra struttura tipica, con cui vengono presentati e venduti questi prodotti, è la conversione di queste analisi in marketing score diversi per ogni consumatore. Tali punteggi si possono basare, ad esempio, sulla probabilità che i consumatori rispondano a delle iniziative di marketing. Ciò permette alle aziende di verificare nelle liste del marketing diretto, chi sono i clienti con basso tasso di risposta.

Per realizzare questi prodotti, vengono impiegati e dati in pasto ai tool analitici, migliaia di dati provenienti sia dai clienti che da fonti pubbliche e commerciali (FTC, 2014).

Infine, l'ultima tipologia di prodotti offerti dai data broker sono i cosiddetti **people search product**. I prodotti "people search" offrono "informazioni sui consumatori ottenute da fonti governative e da altre fonti pubblicamente disponibili, come i siti di social media". Questi prodotti generalmente sono acquistati dai singoli ma possono essere utilizzati anche dalle organizzazioni. Vengono utilizzati per scopi quali "tracciare le attività dei competitor, trovare vecchi amici, ricerca di un potenziale interesse amoroso, indagini personali o per localizzare i registri del tribunale". In generale i prodotti "people search" sono acquistati, da singoli o da aziende per avere accesso a "informazioni personali" sugli individui. Tali prodotti presentano una grande varietà di informazioni tra le quali: "età e data di nascita, numero di telefono, indirizzo di casa, registri di morte, indirizzo e-mail, storia lavorativa, proprietà immobiliare, casellario giudiziale, sentenze di pignoramento o fallimento, informazioni sui social media ecc." (FTC, 2014).

2.2.3 I clienti dei data broker

Di seguito un elenco dei principali clienti dei data broker, divisi per tipologia di prodotto offerto.

	Direct Marketing	Online Marketing	Marketing Analytics	Identity Verification	Fraud Detection	People Search
Alternative Payment Providers[i]				X	X	
Attorneys & Investigators	X					
Automotive Industry	X	X	X			
Consumer Packaged Goods Manufacturers[ii]	X	X	X			
Data Brokers	X	X	X	X	X	
Educational Institutions	X			X	X	
Energy/Utilities	X					
Government Entities	X		X	X	X	X
Hospitality/Travel/ Entertainment	X	X	X			
Individual Consumers						X
Insurance Companies	X		X	X	X	
Lenders/ Financial Services Firms	X	X	X	X	X	X

	Direct Marketing	Online Marketing	Marketing Analytics	Identity Verification	Fraud Detection	People Search
Marketing/ Advertising Firms	X	X	X	X	X	X
Media	X		X			X
Non-profit Entities/ Political Campaigns	X	X		X	X	
Pharmaceutical Firms	X		X			X
Real Estate Services	X				X	X
Retail Companies	X	X	X	X	X	X
Technology Companies[iii]	X	X	X			X
Telecom Companies[iv]	X		X	X	X	

Figura 21: Principali clienti dei data broker divisi per tipologia di prodotto.
Fonte: (FTC, 2014)

Le pratiche di "screening e monitoraggio dei clienti" dei data broker variano a seconda del tipo di prodotto ("advertising o prevenzione delle frodi"), tipo di dati forniti o il tipo di cliente ("istituto finanziario o rivenditore al dettaglio"). Ad esempio, quando lavorano su prodotti "people search" contenenti "informazioni governative e altre informazioni pubblicamente disponibili" fanno uno screening "minimale" dei loro clienti. I clienti dei data broker accedono a questi prodotti attraverso il loro sito web, e in genere non è specificato sul sito l'utilizzo che il cliente farà di questi dati. Tuttavia, i data broker vietano determinati usi dei dati attraverso delle condizioni d'uso presenti nei loro siti web, che possono vietare l'uso dei dati per "scopi illegali". Tali condizioni a volte è richiesto siano accettate dall'utente prima di acquistare il prodotto, a volte invece non è richiesta l'accettazione.

Molti data broker, tuttavia, hanno dei processi accurati e consistenti di "screening e monitoring" dei clienti. Ad esempio, si possono prevedere degli incontri con i clienti o fare delle verifiche sulla conformità legale del business del cliente, come la verifica dell'"indirizzo commerciale". Oppure i data broker possono avere delle policy interne che vietano espressamente di vendere prodotti a clienti facenti parte di determinati settori come "pornografia, investigazione privata, vendita di droghe illegali, vendita di armi". Si può prevedere inoltre un questionario per verificare la "legittimità" del cliente e delle ispezioni del sito del cliente per verificarne la sicurezza.
Nel settore dell'advertising e della mitigazione del rischio, si stipula un contratto scritto che specifica le condizioni d'uso dei dati. Ad esempio, gli usi proibiti possono includere "il riutilizzo e la vendita dei dati senza autorizzazione, la decodifica o il reverse engineering e gli usi in violazione delle linee guida di autoregolamentazione del settore". Per quanto riguarda infine l'accuratezza dei dati, molti data broker dichiarano nelle informative presenti nei loro siti, che i dati dipendono dall'accuratezza delle fonti, scaricando da sé ogni responsabilità. Altri data broker invece, preferiscono tralasciare dichiarazioni sulla qualità dei dati e puntare sulla "qualità predittiva" dei prodotti offerti (FTC, 2014).

2.3 Letteratura sui data broker

La letteratura esistente sui data broker ha cercato di esplorare le caratteristiche del mercato dei dati e il ruolo dei data broker in questo mercato. Un filone della letteratura si concentra sul rapporto tra brokeraggio dei dati e privacy degli individui. Hoofnagle (2004) evidenzia come i data broker utilizzino i dati dei consumatori senza che questi ultimi ne siano pienamente consapevoli o abbiano dato esplicitamente un consenso informato. Federal Trade Commission (2014) sottolinea la poca trasparenza dei data broker che porta a una inevitabile difficoltà a comprendere l'entità del fenomeno, invocando dunque i data broker ad aumentare la trasparenza. Citron e Pasquale (2014), Acquisti, Taylor e Wagman (2016) evidenziano come la raccolta massiva di dati da parte

dei data broker non solo minaccia la privacy ma può portare a segmentare i consumatori in modo discriminatorio ed errato. Shy e Stebancka (2016) analizzano come il surplus dei consumatori ed il benessere collettivo aumentino all'aumentare del livello di protezione della privacy, in un contesto in cui operano un unico data broker monopolista a monte e due imprese a valle. Conitzer et al. (2012), Casadesus- Masanell, Hervas-Drane (2015), Belleflamme et al. (2020), Choi et al. (2019) si occupano delle violazioni della trasparenza e delle attività anticoncorrenziali che possono essere messe in atto, in seguito all'accesso ai dati, da parte di imprese, consumatori o terze parti. Un altro filone della letteratura si concentra sulle strategie dei data broker nella vendita dei dati sui consumatori alle imprese. Francesco Clavorà Braulin e Tommaso Valletti (2016) considerano un data broker che può vendere informazioni a due diverse imprese a valle. Le imprese possono poi utilizzare i dati per applicare prezzi personalizzati. Una impresa è di alta qualità, l'altra è di bassa qualità. Il first best sarebbe raggiunto se le informazioni fossero vendute non esclusivamente, ma in realtà accade che i dati vengono venduti in modo esclusivo ad una sola impresa, è questo sfocia in un'allocazione delle risorse inefficiente. David Bounie et al. (2018) nel loro modello, considerano un data broker monopolista che può scegliere strategicamente la quantità e il tipo di informazioni da vendere alle imprese, le quali useranno i dati per fare discriminazione di prezzo. L'autore conclude che la competizione tra imprese è influenzata dall'ammontare di informazioni fornite dal data broker e in particolare vendendo informazioni parziali si può abbassare il grado di competizione tra imprese. Gu, Madio e Reggiani (2021) considerano due data broker che forniscono informazioni alle imprese a valle ("acquirenti"). Tali data broker possono competere o cooperare tra di loro. In questo articolo gli autori parlano della differenza tra dati "sub-additivi" ovvero il valore dei dati insieme è inferiore rispetto al valore dei dati stand alone, e dati "super-additivi" ovvero il valore dei dati insieme è superiore rispetto al valore dei dati presi separatamente. Gli autori concludono che, se i data broker sono più "efficienti a unire i dataset rispetto alle imprese a valle" è più probabile che ci sia condivisione dei dati tra i due data broker e quindi ci sia una "cooperazione". In questo filone troviamo anche Ichihashi (2020) che si concentra sugli

incentivi dei data broker a condividere i dati e le implicazioni sul mercato, e Montes et al. (2019) che riprendendo Braulin et al. (2016), studiano come cambia il mercato se le informazioni vengono vendute in modo esclusivo ad una sola impresa o viceversa possono essere acquistate e utilizzate contemporaneamente da più imprese.

Capitolo 3: Implicazioni economiche dell'operato dei Data Broker

L’attività dei data broker, tanto pervasiva al punto da influenzare moltissimi settori, quanto opaca, ha messo in evidenza una serie di problematiche sociali di notevole rilevanza. Infatti, se da un lato i consumatori possono trarre beneficio da alcuni dei prodotti offerti dai data broker, ad esempio i prodotti per prevenire le frodi possono aiutare le aziende a smascherare i truffatori che si spacciano per clienti oppure i prodotti people search possono consentire alle persone di ritrovare vecchi amici, compagni di classe o persone con cui hanno perso i contatti, dall’altro lato la raccolta massiccia l’archiviazione e il controllo di dati sensibili degli individui solleva interrogativi profondi sulle implicazioni di privacy e politiche di queste pratiche. Un ulteriore aspetto critico riguarda la scarsa consapevolezza dei consumatori su come i loro dati vengano trattati, molto spesso senza aver dato un esplicito consenso, rendendo di fatto i consumatori impossibilitati a difendersi contro l’uso improprio delle proprie informazioni sensibili. Esistono degli strumenti pratici e dei comportamenti che potrebbero aiutare i consumatori a proteggere i propri dati personali come: l’utilizzo di VPN, utilizzo di estensioni del browser per bloccare i contenuti pubblicitari, disattivare l’accesso alla posizione, ai contatti e ai media quando non necessario oppure configurare le impostazioni di privacy dei social media per limitare il rilascio dei dati. Tuttavia, tali strumenti da soli non bastano in quanto offrono una protezione limitata contro il tracciamento e non riducono significativamente la vulnerabilità dei consumatori nei confronti delle pratiche invasive dei data broker. In questo contesto, dunque, è importante non solo che i cittadini siano consapevoli dell’esistenza di “best practices” che possono aiutare a proteggere i dati, ma è fondamentale che ciò sia accompagnato da adeguati e rigorosi strumenti di legge che tutelino la privacy dei cittadini e garantiscano sanzioni per eventuali violazioni.

Si noti, infatti, che i data broker, per molto tempo, hanno operato in un mercato privo di regolazione, in cui sostanzialmente la raccolta, la vendita e la condivisione di dati, tra

imprese diverse o tra imprese e governi, avveniva senza alcun controllo e senza nessun regolamento che potesse limitare le loro pratiche. Gli scandali emersi recentemente, il più significativo quello di Cambridge Analytica (di cui si discuterà nel Cap. 4), hanno portato gli stati, i governi e i legislatori ad attenzionare maggiormente le questioni relative alla privacy e alla protezione dei dati personali e a interrogarsi se e in quale misura, una regolamentazione più stringente sulla privacy, potesse porre un freno alle pratiche aggressive di raccolta dati dei data broker. Infatti, per tutelare i cittadini da eventuali violazioni della privacy, è stato introdotto in Europa il General Data Protection Regulation (2016), un regolamento che mette la protezione della privacy al primo posto e la considera addirittura come un "diritto fondamentale" dell'individuo. Ad oggi, è lo strumento di legge che ha maggiori effetti sui data broker.

Nonostante i grandi passi avanti fatti in termini di regolazione, i rischi per i consumatori rimangono e sono vari. Infatti, le pratiche dei data broker non intaccano solo la privacy individuale ma posso alimentare forme di discriminazione e avere ripercussioni sulla sicurezza e sul funzionamento democratico. Ad esempio, una profilazione dei consumatori sulla base di informazioni incomplete, può impedire l'accesso all'assistenza sanitaria, o dei report di credito inaccurati possono precludere ai consumatori l'accesso al mercato del credito. Allo stesso modo un individuo potrebbe essere erroneamente classificato come una minaccia per la sicurezza e potrebbe subire restrizioni che ledono la sua libertà personale.

3.1 Il ruolo dei DB nel "Capitalismo di Sorveglianza"

Il Capitalismo di Sorveglianza "rivendica unilateralmente l'esperienza umana come materia prima gratuita da tradurre in dati comportamentali" e ciò ha "profonde implicazioni per l'autonomia individuale e la democrazia, poiché trasforma il regno digitale in uno spazio di controllo e manipolazione piuttosto che di potenziamento". In altre parole, nell'era del capitalismo di sorveglianza, i dati non vengono utilizzati solo per il miglioramento del prodotto o servizi, ma anche come un "surplus

comportamentale" che, dato in pasto ad algoritmi di machine learning, può essere trasformato in "prodotti predittivi che "anticipano ciò che faremo ora e ciò che faremo dopo". Il capitalismo di sorveglianza si fonda dunque sul concetto del "surplus del comportamento", ovvero la pratica dell'accumulazione dei dati basata sull' "estrazione e la monetizzazione del comportamento umano". Grandi compagnie come Google hanno contribuito alla crescita del capitalismo di sorveglianza. Infatti, quando nel 2000 è esplosa la bolla di Internet, l'emergente crisi finanziaria, costrinse Google a trovare altre fonti sostenibili di ricavo. E fu lì che Google si rese conto che "i dati generati dalle interazioni degli utenti potevano essere utilizzati per la pubblicità mirata, dando origine a una risorsa a costo zero pronta per essere monetizzata". I data broker giocano un ruolo da protagonisti nel contesto del capitalismo di sorveglianza e sono attori "pivotali" nel cosiddetto "mercato per i comportamenti futuri" (Zuboff, S., 2019), dato che sono in grado di "dedurre migliaia di attributi da miliardi di consumatori per prevederne il comportamento" (Reviglio, U., 2022). Queste pratiche hanno sollevato dubbi sulla "privacy e la protezione dei consumatori contro discriminazioni ingiustificate e inaspettate", visto e considerato che i consumatori sono "inconsapevoli di come i data broker consolidino, aggreghino, analizzino e vendano i loro dati". Così come logica capitalistica classica il profitto proviene dalla produzione di beni e servizi, realizzati attraverso risorse fisiche e materiali, nel capitalismo di sorveglianza proviene dallo sfruttamento e dalla vendita dei "dati comportamentali degli individui. In questa logica i dati degli individui sono solo delle "commodities", da vendere e scambiare per ottenere un valore aggiunto.

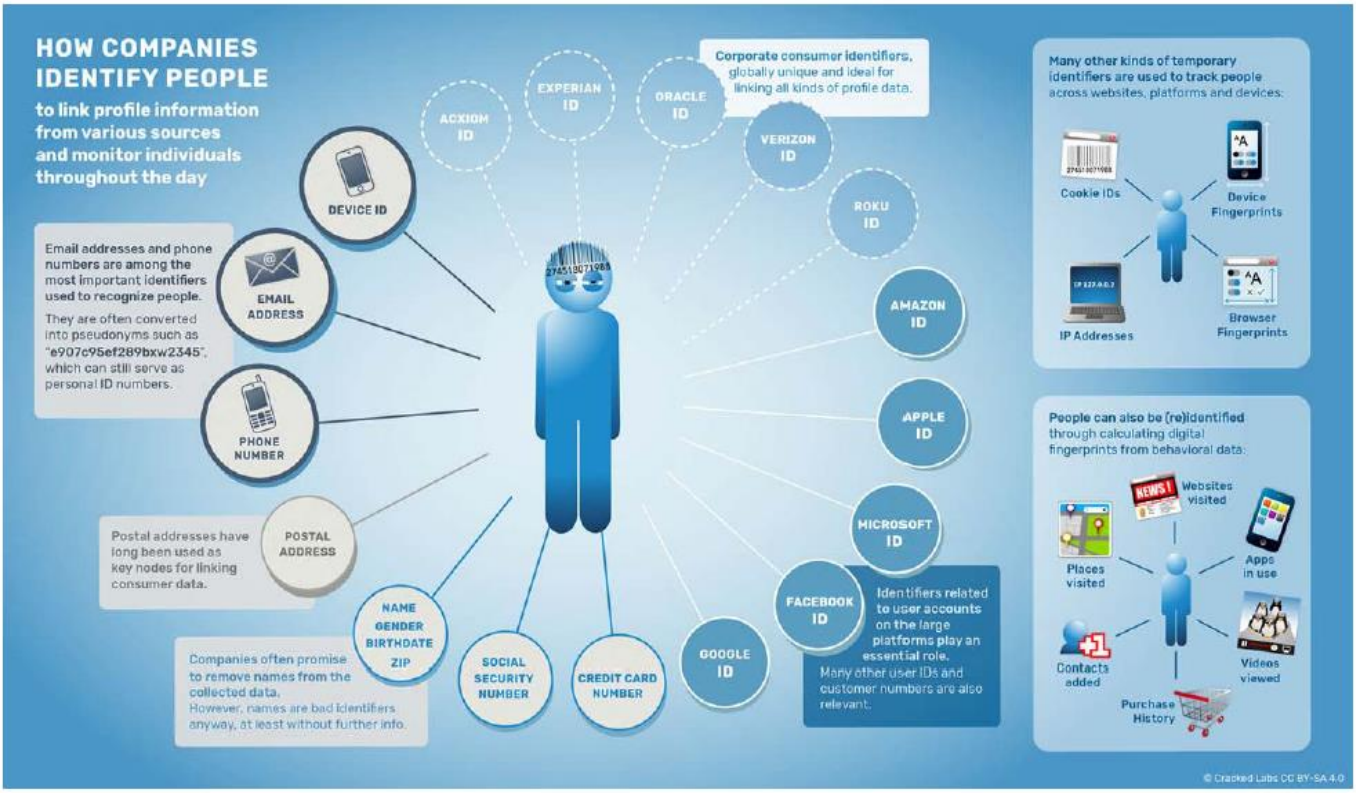

Figura 22: Dati utilizzati dalle imprese per tracciare e profilare i consumatori
Fonte: (Corporate Surveillance In Every Day Life, 2017)

I data broker hanno dunque costruito un ecosistema in cui gli individui sono "costantemente censiti e valutati, indagati ed esaminati, categorizzati e raggruppati, valutati e classificati, numerati e quantificati, inclusi o esclusi e, di conseguenza, trattati in modo diverso" (Christl, W., 2017). Un altro motivo per cui i data broker sono protagonisti nel capitalismo di sorveglianza è che innescano una forma di "arbitraggio informativo", "acquistando, reinterpretando, riconfezionando e vendendo i dati dei consumatori in diversi contesti" e hanno un "ruolo organizzativo, di sfruttamento e pervasivo che non può essere sottovalutato". Tutto ciò contribuisce ad aumentare "l'ineguaglianza nella disponibilità dei dati" e "una privatizzazione dei dati". Infatti, è più semplice acquistare i dati sui consumatori dai data broker piuttosto che raccoglierli da fonti pubbliche. Ad esempio, durante la pandemia di Covid-19, "i passaporti di vaccinazione e i sistemi di identità digitale hanno ampliato le opportunità per gli intermediari di dati di raccogliere dati", con conseguente aumento delle opportunità

d’acquisto per i clienti dei data broker, che rivolgendosi a fonte pubbliche avrebbero riscontrato più difficoltà.

È molto complicato “sfuggire alla sorveglianza” data broker, la cui “acquisizione dei dati è sempre più pervasiva, soprattutto sui social media”. Questo porta con sé diversi problemi, in quanto tra i data broker potrebbe nascere la convinzione che i dati siano “autoesplicativi” e abbiano sempre in pancia “un potere predittivo” anche da soli. Ma le “predizioni” derivano da ipotesi di base, a sua volta prese dalla teoria, che devono necessariamente essere consistenti. Altrimenti si può arrivare facilmente a “inferenze sbagliate e potenzialmente discriminatorie”. Un altro problema è che i dati potrebbero finire nelle mani di “attori malintenzionati”, costituendo dunque una “minaccia alla sicurezza nazionale”.

Per quanto concerne la protezione dei dati, è difficile che ci sia una trasparenza completa, date anche le caratteristiche strutturali del settore. Infatti, si tratta di un approccio “che è subordinato a un discorso di responsabilizzazione dei consumatori, che è stato reso privo di significato nell'ambiente contemporaneo di sorveglianza commerciale pervasiva”. D’altronde, “l'asimmetria della privacy è una pietra miliare del modello di business dei data broker”, poiché si sa pochissimo su di loro e su cosa facciano con i nostri dati. Storicamente le iniziative di trasparenza sono state intraprese con lo scopo di “sostenere i regimi di autoregolamentazione del settore, che hanno ripetutamente fallito nel proteggere la privacy dei consumatori” (Crain, 2018). Questo concetto segue un classico modello in cui la sorveglianza è quasi legittimata dall’ “illusione della scelta” del consumatore, cioè l’illusione che il consumatore abbia il potere di scegliere (accettando o rifiutando i cookie ad esempio). Tuttavia, il rifiuto dei cookie può portare può portare alcune funzioni di un sito a non funzionare correttamente e spesso le informative sulle privacy sono lunghe e prolisse, quindi i consumatori accettano senza sapere a cosa stanno realmente acconsentendo. In sostanza, le aziende danno al consumatore l’illusione di avere il controllo sui propri dati per legittimare la sorveglianza facendola sembrare una scelta consapevole del consumatore (Reviglio, U., 2022).

Molti data broker si tutelano con “accordi di non divulgazione, segreti commerciali e accordi che impediscono agli ex dipendenti di denunciare”. Ciò solleva molti dubbi “sull’efficacia e addirittura sull'applicabilità della regolamentazione dei dati a livello globale” (Reviglio, U., 2022).
I data broker, dunque, hanno cambiato radicalmente l’economia digitale, trasformandola “in uno spazio di controllo e manipolazione piuttosto che di consapevolezza e controllo delle proprie scelte”, con evidenti implicazioni per “l’autonomia individuale e la democrazia” (Zuboff, S.,2019).

3.2 L’attività dei DB e le implicazioni sulla privacy

I data broker possiedono dati sensibili sugli individui, come informazioni demografiche, informazioni relative alla salute, alle preferenze politiche, all’etnia o all’orientamento sessuale. Ad esempio, Acxiom possiede dati su attributi demografici (età, orientamento sessuale, etnia, occupazione), dati finanziari (reddito e patrimonio netto), dati su acquisti e metodi di pagamento (auto, case, uso di carte di credito, digital wallet ecc.), dati sulla salute e addirittura dati geospaziali (geocodifica della latitudine/longitudine) su oltre 2 miliardi e mezzo di consumatori nel mondo, oltre che 225 numeri di telefonici fissi negli Stati Uniti. Verisk invece pubblicizza apertamente oltre 20 miliardi di dati nel settore commerciale e possiede “informazioni sensibili” su oltre 6 milioni di proprietà commerciali, dati sulle frodi assicurative su oltre 1 miliardo di sinistri e “informazioni depersonalizzate” su oltre 1 miliardo e mezzo di conti di credito e di risparmio dei consumatori. Nielsen possiede circa 60,000 segmentazioni dell’audience, che includono dati demografici, online, abitudini d’acquisto, spesa e dati sulla cronologia degli acquisti di oltre 90 milioni di famiglie e dettagli sugli acquisti in oltre 18000 negozi retailers. Possiede inoltre dati sulle transazioni online e offline di compagnie come Visa, Mastercard e American Express (Sherman, 2021). Le pratiche invasive di raccolta dati dei data broker, unite al fatto che si scambiano dati gli uni con gli altri, ha permesso loro di creare un labirintico network di scambio di informazioni, grazie a cui riescono a

ottenere informazioni sempre più puntuali e di qualsiasi tipo sugli individui. Ad esempio, riescono a tracciare i consumatori che si registrano su siti web medici e scoprire se qualcuno di loro è collegato a condizioni sensibili come AIDS o diabete (Kuempel, A., 2016).

Infatti, i data broker possiedono grandi quantità di informazioni sensibili di vario genere. Ad esempio, uno studio di Kim (2023) dal titolo *Data Brokers and the Sale of Americans' Mental Health Data, The Exchange of Our Most Sensitive Data and What It Means for Personal Privacy*, ha rilevato che la sorveglianza dei data broker riesce ad arrivare perfino in aree estremamente sensibili come le informazioni riguardanti la salute mentale degli individui.

Secondo Kim i data broker raccolgono questi dati anche dalle tecnologie "mHealth" (con cui si intende l'uso di dispositivi mobili per la telemedicina, per il monitoraggio da remoto del paziente e per il tracciamento in tempo reale dei parametri vitali di un paziente col fine, ad esempio, di prevenire l'insorgere di una malattia), sfruttando il fatto che tali informazioni non sono tutelate appieno dalla legge HIPAA ("Health Insurance Portability and Accountability Act[14]"), che regolamenta la privacy e la trasparenza solamente per alcune categorie di informazioni mediche. La pandemia di Covid-19 ha fatto registrare un incremento dei problemi di salute mentale, infatti secondo lo U.S Censis Bureau (2021) nel 2020 il 42% degli Americani sotto osservazione ha riscontrato sintomi di ansia, depressione e altri disturbi mentali, con aumento dell'11% rispetto al 2019. In questo contesto è emerso anche un notevole aumento (circa il 200% tra il 2019 e il 2020) del download e dell'utilizzo di applicazioni mHealth, che hanno trovato applicazione nel monitoraggio e nella cura dei problemi di salute psicologica. Kim nel suo studio evidenzia come i data broker riescano ad ottenere i dati sulla salute mentale dalle applicazioni mHealth, approfittando del fatto che alcune piattaforme e alcune categorie di dati non sono sotto la regolamentazione dell'HIPAA e

[14] L' "Health Insurance Portability and Accountability Act" (1996) è una legge statunitense, nata con lo scopo di proteggere la privacy dei dati sanitari dei pazienti. Essa vieta espressamente alle organizzazioni sanitarie di divulgare le informazioni sensibili sulla salute dei pazienti o qualsiasi informazione come nome o indirizzo che possa portare all'identificazione del paziente, senza il loro consenso. Le "informazioni sulla salute" sono definite dalla legge. Fonte: (https://aspe.hhs.gov/report/health-insurance-portability-and-accountability-act-1996)

quindi i titolari di queste piattaforme possono legalmente condividere o vendere i dati con terze parti (inclusi i data broker) senza il consenso degli utenti. Nello specifico, dalla ricerca di Kim è emerso che su 37 data broker contattati, (per mail o per compilazione di modulistica, in cui Kim si era finta una ricercatrice interessata allo studio dei problemi di salute mentale e richiedeva di conoscere l'offerta di dati disponibili in questo ambito) 26 hanno risposto alle domande e 11 di questi erano disposti a vendere i dati sulla salute mentale di cui disponevano. Il prezzo dei dati cambiava in base alla quantità e alla tipologia di dati richiesti, con prezzi che oscillavano da 275 dollari per piccoli dataset, fino ad oltre 100 mila dollari all'anno per l'accesso a database su ampia scala. Inoltre, va sottolineata la facilità con cui sono stati trovati i dati sulla salute; infatti, è stato sufficiente digitare su Google parole come "fornitori di dati sanitari", "dati sulla salute mentale in vendita" o "data broker che vendono dati sulla salute mentale", per trovare aziende disposte a vendere tali dati. Dalla ricerca emerge in modo chiaro un aspetto particolarmente critico che riguarda l'operato dei data broker, ovvero la mancanza di trasparenza sull'uso dei dati, nonché di verifiche approfondite sui compratori. Infatti, tra i 10 data broker più attivi nelle risposte solamente 2, prima di fornire i dati, hanno imposto di firmare degli **accordi di non divulgazione** (conosciuti anche come NDA, ovvero "Non Disclosure Agreement") e hanno richiesto informazioni dettagliate sull'uso previsto dei dati; tuttavia, per quanto concerne la rimanente parte, è emersa una generale assenza di controlli rigorosi sugli acquirenti, nonché sulla gestione dei dati e l'uso dei dati, aspetto che rende estremamente difficile capire come saranno trattati i dati una volta trasferiti ai compratori. Emblematico è il fatto che solamente 1 dei 10 data broker più attivi nelle risposte ha menzionato i "problemi di privacy" e i rischi per gli individui, nel corso delle conversazioni con l'autore dell'articolo. I data broker contattati pubblicizzavano apertamente ed erano disposti a vendere dati sulla salute mentale tra i quali: ansia, depressione, disturbi bipolari, disturbi del sonno, disturbi da deficit di attenzione/iperattività. Un data broker, tra i 10, associava a tali dati anche informazioni su indirizzi postali, razza ed etnia dei soggetti, rendendo i dati "identificabili".

Un'altra problematica che emerge infatti dall'indagine è legato all'insufficiente "**deidentificazione**" dei dati. Sebbene la maggior parte dei data broker contattati sostiene di anonimizzare i dati prima di rilasciarli, non sempre questa operazione è sufficiente a garantire l'anonimato. Infatti, se la deidentificazione del dato non è sufficiente, incrociando quei dati con le altre fonti disponibili, è possibile ricomporre l'identità dell'individuo. Tutto ciò solleva pesanti timori legati al fatto che, individui con una precaria salute mentale, possono essere identificati, contattati ed esposti a rischi di discriminazioni, abusi, tentativi di frode, violazioni dei diritti (ad esempio un ente assicurativo potrebbe negare l'apertura di una polizza o aumentare il premio).

Ciò che si evince dallo studio è la quasi totale mancanza di trasparenza nel mercato dei dati sulla salute mentale, dovuta principalmente all'assenza di una regolamentazione federale sulla privacy che tuteli esplicitamente questo genere di dati. In questo contesto di opacità, i data broker possono infatti continuare a raccogliere e vendere i dati senza alcuna limitazione, trasparenza e senza alcun controllo, con conseguenze ancor più pericolose per le persone più vulnerabili come gli individui affetti da patologie psicologiche. Infine, l'autore suggerisce due possibili soluzioni per limitare il problema: l'introduzione di una legislazione federale sulla privacy che tuteli esplicitamente la vendita dei dati sensibili (inclusi quelli sulla salute mentale) oppure l'estensione della copertura legale dell'HIPAA anche alle tecnologie emergenti mHealth.

Un altro rischio è legato al fatto che i dati sensibili sono a forte rischio di **disclosure non autorizzata**, in quanto i data broker, ad esempio, possono subire un data breach o degli accessi illegittimi ai loro database. In quest'ultimo caso è emblematico ciò che è accaduto al data broker Choice Point.

ChoicePoint era uno dei maggiori e più profittevoli data broker in America, nato nel 1997 come spin-off del rinomato credit bureau Equifax. Il suo core business era rappresentato dalla fornitura di report di credito agli assicuratori, ma forniva anche prodotti e servizi sofisticati per il pubblico e per le forze dell'ordine (Otto, P. et al., 2007). Ad esempio, deteneva contratti multimilionari col "Department of Homeland Security" (dipartimento federale responsabile della pubblica sicurezza) e col

“Department of Justice”, a fronte della gestione di siti web appositi per facilitare ed accelerare la ricerca dei potenziali sospettati da parte dei dipartimenti federali. Il suo business era cresciuto rapidamente, tanto da passare da un fatturato di 585 milioni di dollari nel 2000 fino a circa 1 miliardo nel 2006 (Stevens, G., 2007). Le fonti di raccolta dei dati erano variegate, ma prevalentemente pubbliche, tanto da arrivare a collezionare circa 19 miliardi di record da fonti pubblicamente disponibili (Otto, P. et al., 2007).

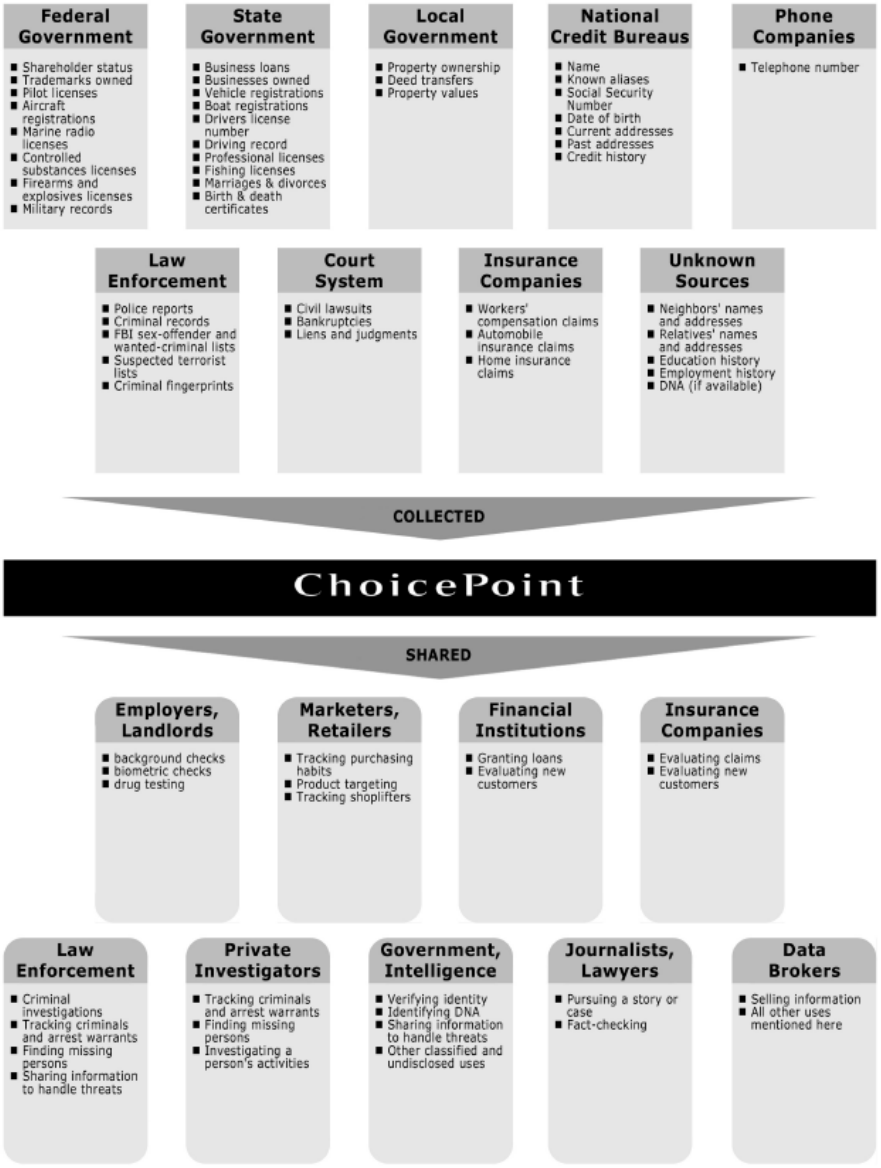

Figura 23: Flusso di raccolta e condivisione dati di ChoicePoint.
Fonte: (Otto, P. et al., 2007)

Per policy aziendale, ChoicePoint garantiva l'accesso ai propri database solo dopo la presentazione da parte del cliente di documenti, (attestato di iscrizione al registro delle imprese, registri aziendali collocati presso le agenzie governative, patenti di guida ecc.), che permettessero di poter verificare l'identità del cliente e la reale esistenza dell'azienda (Stevens, G., 2007). Tuttavia, nel 2003, un gruppo di truffatori riuscì a superare queste verifiche utilizzando delle licenze commerciali false, ottenute utilizzando nomi, numeri di telefono, numeri di previdenza sociale (SSN) e indirizzi, corrispondenti a persone reali e di cui si erano impossessati precedentemente. ChoicePoint attirò a sé aspre critiche poiché non segnalò immediatamente il fatto a tutti coloro i cui dati personali erano stati trafugati, ma solamente ai residenti in California, in quanto era l'unico Stato ad avere una legge apposita in materia di trasparenza sulla violazione dei dati personali (il "Security Breach Information Act" (2003) impone infatti "a qualsiasi organizzazione operante in California di divulgare le violazioni dei dati ai residenti in California quando si verifica un accesso non autorizzato a informazioni personali non criptate"). Tuttavia, dopo una forte tempesta mediatica e pubblica, notificò un avviso anche alle vittime non residenti in California. Sul finire del 2005, ChoicePoint notificò a circa 163 mila persone che le loro informazioni personali erano state compromesse da "parti fraudolente", le quali hanno effettuato circa 17 mila ricerche nei database di ChoicePoint. In seguito, le indagini porteranno ad accertare il furto di 800 identità appartenenti a individui reali, le cui informazioni sono state utilizzate per accedere ai database di ChoicePoint. L'utilizzo di identità (rubate) appartenenti a persone reali e senza alcun precedente penale, ha permesso ai truffatori di non essere individuati, nonostante le verifiche sull'identità dei clienti effettuate di routine da ChoicePoint (Otto, P. et al., 2007). Nello stesso anno la Federal Trade Commission avviò un'indagine nei confronti di ChoicePoint, accusando l'azienda di non aver implementato procedure adeguate per la verifica sull'identità dei propri clienti e di aver fornito dunque l'accesso alle informazioni sensibili (tra cui informazioni creditizie) all'interno del proprio database a individui con credenziali dubbie o inventate e che non avevano alcuno ragione lecita per detenere ed utilizzare quelle informazioni, in esplicita

violazione del Fair Credit Reporting Act. L'indagine si concluse nel 2006, con ChoicePoint che, per risolvere le accuse (infrazione del Fair Credit Reporting Act) mosse dalla FTC, accettò di pagare una multa pari a 10 milioni di dollari e di pagare altri 5 milioni di dollari come risarcimento ai consumatori danneggiati dalla violazione. Inoltre, l'accordo con la FTC prevedeva anche l'adozione di: misure di sicurezza più stringenti per i dati sui consumatori, procedure nuove che garantissero che i report sui consumatori fossero forniti solo ad aziende operanti nel perimetro della legge e di finanziare e sottoporsi a verifiche periodiche (ogni due anni) da parte di revisori indipendenti fino al 2026 (Stevens, G., 2007).

Nel 2009 l'azienda è stata acquistata da LexisNexis ed è stata rinominata "LexisNexis Risk Solutions".

Uno scandalo simile a quello di ChoicePoint è stato quello del data broker LexisNexis. Nel 2004 LexisNexis, sussidiaria di Reed Elsevier (adesso nota come RELX PLC), acquistò il data broker Seisint, il quale si occupava anche di fornire dati per un progetto denominato Matrix, ovvero un database nato con lo scopo di aiutare i governi federali a rintracciare criminali e terroristi e finanziato fino a poco tempo fa dal governo americano. Nel 2005 LexisNexis comunicò che ignoti non autorizzati erano riusciti ad ottenere l'accesso alle informazioni personali di circa 32 mila clienti. Tuttavia, dopo poche settimane il numero dei clienti è salito a 310 mila. Questi individui erano riusciti a violare il sistema utilizzando nomi utenti e password di clienti reali della società. Nello specifico gli hacker hanno in qualche modo ottenuto le password degli utenti paganti del servizio Acurint di Seisint (unità operativa di LexisNexis), il quale di solito forniva pacchetti di informazioni su singole persone dietro il corrispettivo di 4,50 dollari ciascuno. Grazie a queste password, gli hacker hanno avuto accesso a informazioni come nomi, indirizzi di residenza, numeri di previdenza sociale e numeri di patente di guida (Stevens, G., 2007). La FTC, in seguito, avviò un'indagine nei confronti di Reed Elsevier (società capogruppo) e Seisint (partecipata di LexisNexis) per violazione delle normative sulla sicurezza dei dati personali. Le due società, per risolvere le accuse,

hanno accettato l'ordine della FTC (2008)[15] che prevedeva l'implementazione di programmi minuziosi di sicurezza informatica a tutela della privacy e dell'integrità dei dati personali degli individui. Questi programmi prevedevano: identificazione dei fattori di rischio, controlli tecnici e amministrativi a tutela delle informazioni personali, valutazioni indipendenti (da fare ogni due anni) da parte di terzi che attestino la conformità, all'ordine, delle procedure adottate per tutelare la privacy dei consumatori, di conservare i registri e i rapporti sulla conformità per una determinato periodo di tempo e di notificare alla FTC qualsiasi variazione nella struttura aziendale che potrebbe condizionare la conformità all'ordine, fino al 2028.

Un altro caso di violazione della privacy è quello di un data broker di nome "Social Data", che nel 2020, a causa di un errore di configurazione del cloud, aveva lasciato un database che si affacciava su Internet senza alcuna protezione e senza alcuna password di accesso, rendendolo di fatto accessibile a chiunque (spesso i data broker non investono abbastanza in cyber security). Tale database conteneva dati sui social media (nomi, link a siti web personali o aziendali, immagini, video e indirizzi e-mail) di 235 milioni di persone (Ikeda, S., 2020).

Oltre ai rischi riguardanti la privacy dei cittadini, ricordiamo un altro rischio connesso alla privacy e alla condivisione dei dati sensibili, ovvero l'**accuratezza** dei dati e dei report da loro condivisi. Secondo le stime, tali report vengono acquistati da vari clienti quali team di recruiter, assicuratori, istituti finanziari, enti pubblici ecc. Se tali report contengono errori, e ad esempio collegano erroneamente il consumatore a reati penali o crimini, possono causare discriminazione ingiustificata che si può tradurre in una mancata assunzione o in un rifiuto per un alloggio residenziale pubblico o un appartamento privato (Sherman, 2021). Un caso di cronaca riguarda il data broker Spokeo, che ricevette una multa dalla FTC pari a 800,000 dollari, per aver commercializzato degli strumenti di screening e selezione del personale, per datori di lavoro e recruiter, basati su profili imprecisi dei consumatori. Oltre all'imprecisione dei

[15] https://www.ftc.gov/sites/default/files/documents/cases/2008/08/080801reeddo.pdf

profili, la FTC sosteneva che Spokeo non si è assicurato che le informazioni vendute fossero utilizzate solo per scopi legali, non ha richiesto il consenso al trattamento dei dati e non ha informato il consumatore del suo diritto a chiedere al data broker un resoconto delle informazioni possedute su di lui, tutte pratiche in violazione del Fair Credit Reporting Act (1970) (FTC, 2012).

L'accuratezza dei report sui consumatori detenuti dai data broker è stata analizzata da Venkatadri, G. et al. (2019) in uno studio dal nome *Auditing Offline Data Brokers via Facebook's Advertising Platform*. In questo studio i ricercatori si sono occupati di analizzare la portata e l'accuratezza dei dati raccolti (offline) da alcuni dei più importanti data broker come: Acxiom, Experian, Epsilon e Oracle Data Cloud (nota come "Datalogix", prima di essere acquisita da Oracle). Per questa analisi i ricercatori hanno utilizzato la piattaforma pubblicitaria di Facebook, la quale, oltre a consentire agli inserzionisti di acquistare spazi pubblicitari e pubblicare annunci mirati (sfruttando le informazioni raccolte da Facebook attraverso le interazioni degli utenti), collabora anche con i data broker sfruttando le informazioni (offline) in loro possesso per ampliare i propri profili sugli utenti. L'unione delle informazioni online (raccolte da Facebook) e offline (raccolte dai data broker) permette a Facebook di offrire opzioni di targeting degli utenti più precise, le quali vengono sfruttate dagli inserzionisti per migliorare l'efficienza delle campagne di pubblicità mirata. I ricercatori hanno utilizzato la tecnica dei Treads ("Transparency-Enhancing Advertisements), ovvero annunci pubblicitari mirati che consentono all'inserzionista di rivelare agli utenti quali informazioni sul loro conto sono state utilizzate per il targeting, inserendo ad esempio tali informazioni nell'annuncio stesso. Questa tecnica permette di aumentare la trasparenza sull'utilizzo dei dati da parte delle piattaforme di advertising (come quelle di Facebook e Google).

In questo studio, i ricercatori hanno inviato a 183 utenti annunci contenenti informazioni (inclusi quelle tipicamente detenute dai data broker) che potevano potenzialmente essere attribuite a quegli specifici utenti. Un primo annuncio è stato inviato a tutti i partecipanti allo studio in modo tale da capire quanti di loro fossero "targettizzabili" (gli annunci vengono indirizzati agli utenti sulla base di criteri di targeting (attributi) scelti

dall'inserzionista; quindi, se un utente riceve l'annuncio significa che soddisfa questi criteri e fa parte del pubblico potenziale per la campagna pubblicitaria), un secondo annuncio invece, è stato indirizzato solo agli utenti "targettizzabili" che possiedono un particolare attributo rivelato dai ricercatori. Se un utente riceve entrambi gli annunci significa che quell'utente possiede l'attributo specifico a lui associato, permettendo così ai ricercatori di verificare quali informazioni fornite dai data broker sono associate agli utenti all'interno della piattaforma pubblicitaria di Facebook.

I risultati della ricerca hanno messo in evidenza come il 90% degli account Facebook negli Stati Uniti sia associato a informazioni fornite dai data broker, permettendo dunque alla suddetta piattaforma pubblicitaria un'estesa combinazione di dati online e offline da usare per il targeting pubblicitario. Tuttavia, chiedendo un feedback ai 183 volontari sull'accuratezza dei dati a loro attribuiti dai data broker, è emerso che il 40% degli attributi a loro assegnati erano del tutto imprecisi.

	Data Brokers	Acxiom	Datalogix	Experian	Epsilon	Others
Responses	1,432	95	728	55	18	536
Not at all accurate	40.5% ± 2.5%	27.4% ± 9.0%	42.0% ± 3.6%	21.8% ± 10.9%	55.6% ± 23.0%	42.2% ± 4.2%
Somewhat accurate	13.6% ± 1.8%	14.7% ± 7.1%	14.4% ± 2.6%	20.0% ± 10.6%	0.0% ± 0.0%	12.1% ± 2.8%
Mostly accurate	5.2% ± 1.2%	4.2% ± 4.0%	6.2% ± 1.7%	12.7% ± 8.8%	5.6% ± 10.6%	3.4% ± 1.5%
Completely accurate	40.6% ± 2.5%	53.7% ± 10.0%	37.4% ± 3.5%	45.5% ± 13.2%	38.9% ± 22.5%	42.4% ± 4.2%

Figura 24: Accuratezza degli attributi sugli utenti provenienti da diversi data broker. La prima colonna contiene l'accuratezza calcolata considerando tutti gli attributi insieme.
Fonte: (https://hal.science/hal-02069470/file/databrokers-measurement_finalCameraReady.pdf)

Anche dati sensibili come le informazioni finanziarie (reddito, patrimonio netto e investimenti) mostravano un alto livello di imprecisione, la percentuale in quest'ultimo caso eri pari addirittura al 47% per reddito e patrimonio netto, e il 56% per gli investimenti.

Un aspetto che emerge quasi costantemente è che le categorie più a rischio di di report inaccurati sono quelle "già di per sé oppresse o emarginate". Ad esempio, uno studio di Kaplan, L et al. (2022) indica che "le minoranze etniche hanno maggiori probabilità di avere informazioni errate nei loro rapporti di credito rispetto agli individui bianchi o a quelli che vivono in località più ricche".

3.3 Le implicazioni sui diritti civili, la sicurezza e sul funzionamento democratico

I dati sensibili, collezionati dai data broker, non rimangono "dati grezzi", ma subiscono delle elaborazioni funzionali alla successiva vendita o condivisione. Infatti, dopo aver raccolto i dati grezzi da varie fonti, tali dati vengono uniti, per creare un "composito dettagliato" del consumatore. Questa pratica è chiamata "effetto di aggregazione". In seguito, i data broker, usano algoritmi di apprendimento per fare inferenze sui consumatori, che vengono inseriti in categorie relative "all'etnia, al reddito e all'appartenenza politica" (Kuempel, A., 2016).

Queste "composizioni dettagliate", secondo la Federal Trade Commission (2014), possono "facilitare l'invio di pubblicità mirate" su salute, etnia, religione o prodotti finanziari, che alcuni consumatori potrebbero trovare **discriminatorie** e che possono "minare la loro fiducia nel mercato". Infatti, nei database di molti data broker si trovano categorie che discriminano i consumatori su base etnica e sulla stabilità economica, come "Urban Scramble" o "Mobile Mixers". Tali categorie descrivono consumatori afroamericani e latini con basso reddito, e possono essere utilizzate dalle banche per proporre prestiti ad alti tassi di interesse o prodotti finanziari ad alto rischio, facendo leva sull'esigenza di "denaro rapido" di questa fascia della popolazione (Kuempel, A., 2016). Tra altre categorie nello stesso ambito ricordiamo: "Hard Times" che descrive anziani single tra i 50 e i 75 anni, tipicamente afroamericani ma anche ispanici ed asiatici, "che vivono in quartieri poveri della città e sono generalmente privi di sostegno familiare", "Tradizioni senza tempo", che include immigrati, alcuni dei quali in età pensionabile, che parlano male l'inglese e preferiscono lo spagnolo e che hanno redditi inferiori alla media" , "Lavoro e cause" (consumatori tra i 40 e i 50 anni con redditi inferiori rispetto alla media che vivono in case con più unità abitative", "Genitori metropolitani" (genitori single che hanno generalmente un diploma di scuola superiore o di scuola professionale) e infine "Etnici che lottano per la seconda città", "Rurali che ce la fanno a malapena" e "X-tra Needy" (FTC, 2014),(Committee On Commerce,

Science, And Transportation Of U.S Senate, 2013). Queste segmentazioni su base etnica e finanziaria possono consentire alle imprese che acquisteranno queste liste a sottoporre i consumatori a prezzi differenziati. Ad esempio, possono essere utilizzate per individuare consumatori con limitato accesso alle banche, tipicamente nuovi immigrati e neolaureati, e inviare pubblicità su "prestiti subprime" (ad alti interessi). Queste categorie vengono generalmente chiamate "Financially Challenged" o "Underbanked" (Kuempel, A., 2016).

Altre categorie che, insieme alla vulnerabilità finanziaria, sottolineano l'età avanzata dei consumatori sono: "Rurale perenne", che include uomini e donne single di età superiore ai 66 anni con scarso livello di istruzione e bassa ricchezza oppure "Anziani parsimoniosi", che comprende single tra i 60 e i 70 anni che si trovano in fasce di reddito inferiori alla media. Questa popolazione "finanziariamente fragile" e in età avanzata, è fortemente a rischio di **discriminazione** o **frode** (Committee On Commerce, Science, And Transportation Of U.S Senate, 2013).

Figura 25: Categorie utilizzate per identificare la popolazione finanziariamente fragile.
Fonte: (Committee On Commerce, Science, And Transportation Of U.S Senate, 2013)

Un esempio emblematico frode, connessa ai dati posseduti dai data broker, è il caso della compagnia InfoUSA, un data broker che aveva dati su oltre 210 milioni di americani. InfoUSA nel 2007 pubblicizzava e vendeva, a individui e altre compagnie, liste contenenti nomi di consumatori anziani. Questi nomi venivano poi utilizzati dagli acquirenti per proporre vendite fraudolente agli anziani. InfoUSa pubblicizzava liste come "Anziani in cerca di opportunità", cioè persone "alla ricerca di un modo per fare soldi", "Anziani sofferenti", persone col cancro o affette da Alzheimer, "Vecchi ma buoni" persone sopra i 55 anni che amavano il gioco d'azzardo. Le liste venivano pubblicizzate con frasi come "queste persone sono dei creduloni, vogliono credere che la loro fortuna possa cambiare". Gli anziani venivano poi contattati e invitati con l'inganno a rivelare le loro informazioni bancarie. Poi questi soggetti andavano in banca e svuotavano il conto corrente delle persone contattate. Generalmente contattavano reduci dalla Seconda Guerra Mondiale e insegnanti in pensione, sostenendo di essere dei funzionari governativi, assicurativi e di aver bisogno dei loro dati personali per aggiornare dei fascicoli. Una delle vittime è stato il veterano di guerra novantaduenne

Richard Guthrie. Il suo nome era comparso nei database di InfoUSA, in una lista chiamata "Astroluck", in quanto aveva partecipato in passato a delle lotterie. I ladri si erano spacciati per impiegati statali e sostenevano che i computer della Social Security Administration (l'istituto di Previdenza Sociale americano) erano in tilt e i registri delle prescrizioni mediche erano incompleti. Quindi riferivano che le pensioni e le prescrizioni gratuite sarebbero arrivate in ritardo a meno che non avesse fornito le sue informazioni bancarie. Richard Guthrie spaventato da questa affermazione, ha rivelato le sue informazioni bancarie permettendo così ai ladri di andare in banca e rubare i suoi risparmi personali che ammontavano a 100 mila dollari (Duhigg, C.,2007). Questo caso da un'idea dei notevoli rischi, di pratiche fraudolente (specialmente per le fasce più vulnerabili della popolazione), connesse all'archiviazione, da parte dei data broker, di dati sensibili.

I data broker possono utilizzare i dati sensibili su razza o etnia per pubblicizzare prodotti o servizi in modo discriminatorio, a scapito dei **diritti civili**. In tal senso, è significativo il caso di Facebook del 2016.

La piattaforma, in quell'anno, consentì agli inserzionisti di pubblicare annunci sulla piattaforma che escludevano determinati individui su base etnica e razziale. Gli inserzionisti, dunque, con la complicità di Facebook, potevano utilizzare la piattaforma per pubblicare annunci immobiliari che escludevano esplicitamente dalla visualizzazione etnie specifiche come afroamericani, ispanici e asiatici.

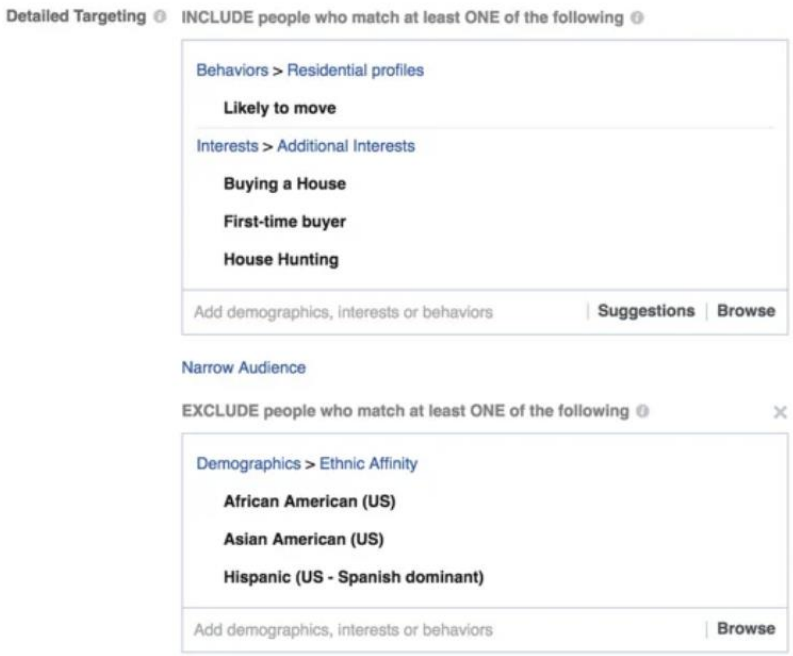

Figura 26: Dettagli di un annuncio pubblicitario su Facebook con targeting discriminatorio.
Fonte: (https://www.propublica.org/article/facebook-lets-advertisers-exclude-users-by-race)

Facebook tramite la voce di Steve Satterfield, responsabile delle privacy e delle politiche pubbliche, si difese dicendo che le loro policy proibiscono di pubblicare annunci con target discriminatori e che si è trattato solo di una campagna di misurazione delle prestazioni del marketing, pratica molto comune a suo dire tra gli inserzionisti, che così possono ad esempio "eseguire una campagna in inglese che escluda il gruppo di affinità ispanica per vedere quanto bene la campagna si comporta rispetto all'esecuzione di quella campagna pubblicitaria in spagnolo" (Angwin, J., Parris Jr, T., 2016). Il caso si risolse infine con un nulla di fatto e a Facebook non fu imposta alcuna sanzione. Tuttavia nel 2022 la piattaforma, dopo un accordo col Dipartimento di Giustizia, ha accettato di eliminare dalla sua piattaforma pubblicitaria il tool che consente agli inserzionisti di escludere dalla visualizzazione degli annunci determinate persone in base all'etnia o alla razza (Tobin, A., Kofman, A., 2022).

Un altro caso simile riguarda la Toyota Motor Credit Corporation. Infatti, nel 2016, la Toyota dopo aver raggiunto un accordo col Dipartimento di Giustizia Americano e Consumer Financial Protection Bureau , ha risarcito per una cifra pari a 21,9 milioni di dollari, i richiedenti prestito tra 2011 e il 2016, in quanto aveva applicato tassi di

interesse più alti nei confronti di afroamericani e asiatici rispetto a bianchi non ispanici a parità di credit score. Infatti, la discriminazione era soltanto etnica e non riguardava criteri sulla solvibilità del richiedente prestito o sulla sua rischiosità. Ciò è stato possibile grazie alla disponibilità di Toyota di dati demografici acquisiti da fonti pubbliche e private, tra cui i data broker, che le hanno permesso di identificare i consumatori e applicare tassi di interesse differenti in base all'etnia. Addirittura, gli afroamericani erano costretti a pagare 200 dollari in più per la durata del loro prestito, mentre gli asiatici 100 dollari in più. Dopo il fatto, Toyoya si impegnò col Dipartimento di Giustizia a promuovere un accesso equo ai finanziamenti e che facesse leva solo sul merito creditizio e non su aspetti, non finanziari, come etnia o razza (U.S Department of Justice, 2016).

I report dei data broker hanno molto mercato tra i proprietari di immobili, infatti secondo Kirchner (2020), 9 proprietari di immobili su 10, si rivolgono ai data broker, chiedendo dei report sulla storia giudiziaria degli acquirenti o affittuari, o acquistando "prodotti people search" per cercare informazioni su di loro, ai fini di garantire la sicurezza della loro proprietà. Tra le aziende che offrono queste tipologie di report e prodotti ci sono RealPage e Rentgrow. Per avere un'idea dei rischi per i diritti civili connessi alla vendita di questi report, basti pensare a quanto raccontato da Kirchner (2020), ovvero il caso di una studentessa di giurisprudenza, che nel 2015 fece domanda per un alloggio in affitto e tale domanda fu respinta dal proprietario, perché in dei report della società RealPage, il suo nome era collegato ad un reato penale commesso anni prima. Tuttavia, nel report non compariva la fattispecie del reato, quindi il proprietario diede per scontato che fosse un reato grave. Alla fine, tale reato si scoprirà essere una multa per eccesso di velocità risalente al 2011. Alla fine, la ragazza fece causa a RealPage per diffamazione, ma la causa fu archiviata e si risolse tutto in un nulla di fatto.

In modo simile le compagnie assicurative (assicurazioni sanitarie e sulla vita) acquistano report (contenenti dati su razza ed etnia) dai broker di dati per stimare il costo del servizio. Tali report vengono preparati tramite tool predittivi che, se basati su dati inaccurati o discriminatori, possano prevedere costi lievitati soprattutto per le

minoranze. Ciò rappresenta una minaccia per i diritti civili, poichè dati gli alti costi dei servizi (dovuti a informazioni discriminatorie o non pertinenti, come abitudini televisive, gusti musicali o quartiere di residenza e non a criteri oggettivi, ad esempio un costo del servizio proporzionale ai problemi di salute del cliente), potrebbero portare le minoranze a non riuscire a sostenere le spese per l'assicurazione sanitaria. Secondo un'indagine di Allen, M. (2018) grandi imprese nel campo dell'assistenza sanitaria, come Optum e IBM Watson Health che possiedono milioni di dati di consumatori americani su "diagnosi mediche, test, prescrizioni e dati socieconomici", uniscono i propri dati con i report acquistati dai data broker, per prevedere il costo dell'assistenza sanitaria. E costi elevati di assistenza sanitaria potrebbero portare le compagnie ad aumentare il premio assicurativo, per riuscire a coprire i costi delle spese mediche sostenute dai pazienti. I data broker inseriscono dunque dati relativi a etnia, stato civile, quartiere di residenza in complicati algoritmi che prevedono il potenziale costo dell'assicurazione sanitaria. Per i data broker, ad esempio, se una persona fa parte di una minoranza etnica è probabile che viva in un quartiere pericoloso e degradato, l'algoritmo traduce ciò in maggiori rischi per la salute e dunque maggiori costi dell'assicurazione. LexisNexis, ad esempio, ha affermato di utilizzare "442 attributi personali non medici per prevedere i costi medici di una persona". Tali dati riguardavano etnia, precedenti penali e sicurezza del quartiere. McCulley, direttore delle soluzioni strategiche di LexisNexis ha confermato che le previsioni fatte dagli algoritmi si basano su dati personali, ma ha ribadito che le previsioni e i punteggi ottenuti servono per aiutare i pazienti a ottenere le cure di cui hanno bisogno e non a prevedere quanto pagherebbero per l'assicurazione sanitaria. Tuttavia, l'utilizzo di questi dati, specialmente se inaccurati, potrebbe favorire le disuguaglianze nella misura in cui "potrebbe portare a far pagare di più le persone povere, rendendo difficili le cure di cui hanno bisogno, o potrebbe portare i datori di lavoro a decidere di non assumere persone con dati che potrebbero indicare costi medici elevati in futuro". Le aziende, dunque, dovrebbero valutare l'accuratezza dei dati e in che modo questi dati potrebbero discriminare i poveri e le minoranze (O'Neil, C., 2016).

Un altro rischio riguardante i report dei data broker, è quello che questi vengano acquistati e utilizzati col mero scopo di **discriminare** gli altri, senza alcun motivo apparente. Ad esempio, il sito ultracattolico "The Pillar" nel 2021, ha scoperto e reso noto, condividendo anche l'informazione con altri vescovi cattolici, il nome di un sacerdote gay. Per fare ciò erano stati acquistati dai data broker, i dati sull'utilizzo di Grindr dell'individuo. Tali dati includevano dati sulla localizzazione e ricerche sull'app (NBC News, 2021).

Le ca

Un'altra minaccia per i diritti civili riguarda il fatto che, le assicurazioni, potrebbero utilizzare delle segmentazioni "innocue" acquisite dai data broker in modi che suscitano perplessità. Ad esempio, un data broker potrebbe dedurre che un consumatore è appassionato di moto e inserirlo in una segmentazione chiamata "Appassionati di motociclismo". Una compagnia assicurativa che ha acquisito la segmentazione potrebbe dedurre che il consumatore ha un comportamento rischioso e applicare dunque un costo alto per la polizza assicurativa. Allo stesso modo un segmento contenente persone con "Interessi per il diabete" potrebbe portare la compagnia assicurativa a classificare il cliente come ad alto rischio aumentando il premio assicurativo (FTC, 2014).

Sia in Europa che negli Stati Uniti, le forze dell'ordine si affidano ai tool di profilazione e analisi forniti dai data broker, in aggiunta agli strumenti a loro disposizione, per localizzare facilmente i sospettati o fare ricerche su di loro. Ad esempio, negli Stati Uniti è molto ricercato un prodotto people search chiamato "Accurint for Law Enforcement Plus" fornito da LexisNexis ed utilizzato da più di 4.000 agenzie federali, statali e locali delle forze dell'ordine. Tale strumento permette alle forze dell'ordine di "localizzare facilmente sospetti, testimoni e fuggitivi", "scoprire rapidamente beni" e "scoprire collegamenti tra persone, aziende, beni e luoghi". Oppure TransUnion pubblicizza un prodotto simile chiamato "TLOxp for Law Enforcement", utilizzato da più di 100,000 ufficiali delle forze dell'ordine in tutto il paese, che offre un vasto database online contenente registri pubblici e privati, con informazioni su "persone, aziende, assets e posizioni" supportato da un database di "trilioni di registrazioni". Tali registri

contengono anche dati privati come “e-mail, social media, avvistamenti di veicoli e informazioni sulla posizione” per creare “strumenti di ricerca e mappatura personalizzati per la polizia”. Tuttavia, quando le forze di polizia si affidano eccessivamente a questi “prodotti di ricerca” dei data broker, anche piccole imprecisioni nei dati possono portare a gravi violazioni dei **diritti civili**. Ad esempio, un errore o uno scambio di identità potrebbero portare ad un arresto ingiustificato o nel peggiore dei casi ad uno eccessivo della forza da parte della polizia contro la persona sbagliata. Gli errori si possono manifestare in diversi modi, ad esempio si possono scambiare accidentalmente i record di persone che hanno lo stesso nome oppure durante la digitalizzazione dei documenti, il programma di riconoscimento automatico dei caratteri potrebbe riconoscere in modo errato un ID numerico, collegandolo ad un individuo diverso. Inoltre, i report possono diventare obsoleti e mancare di dati aggiornati. Ad esempio, un individuo potrebbe essere stato accusato di un reato e in seguito assolto, ma nei report dei data broker compare solo l’accusa e non la disposizione di non colpevolezza, creando così un profilo profondamente impreciso che può portare a conseguenze gravi, come compromettere la reputazione di una persona o sfociare in forme di sorveglianza mirata senza giustificazione e altre forme che potrebbero a violare la privacy di un individuo o a compromettere la sua libertà personale. Oppure un report contenente dati incompleti, non aggiornati oppure contenente per errore dei reati penali a carico di un individuo, qualora finisse nelle mani dell’ICE (“United States Immigration and Customs Enforcement) potrebbe implicare per un cittadino immigrato, l’espulsione o la “removibilità” dello status di immigrato (Rieke, A. et al., 2016). I data broker, con i loro prodotti di ricerca e algoritmi di profilazione, contribuiscono ad alimentare la pratica della “profilazione razziale”, ovvero l’uso di dati sull’etnia, razza e religione, da parte della polizia, per perseguitare e prendere di mira individui. La profilazione avviene associando caratteristiche come etnia e luogo d’origine a una fattispecie di reato, ritenendo di fatto tale reato come tipico di persone con quelle determinate caratteristiche demografiche (American Civil Liberties Union, 2005).

Esistono anche dei prodotti di nicchia specificatamente progettati per l'uso da parte della polizia. Ad esempio, è famoso un prodotto chiamato "Bluejay" che viene pubblicizzato con la dicitura "law enforcement Twitter crime Scanner", che consentono alla polizia di monitorare facilmente "utenti, parole chiave o aree geografiche selezionate, come il luogo di una protesta". Tali prodotti seguono i comportamenti sui social media, come ad esempio i tweet pubblicati e forniscono alla polizia un modo a basso costo per sorvegliare le persone (Rieke A. et al., 2016). Tuttavia, possono essere usati in modo improprio. Per esempio, la polizia di New York, tra il 2013 e il 2014 ha utilizzato tali prodotti, acquistandoli dai data broker, per sorvegliare ragazzi di Harlem in età scolastica, compresi ragazzi senza precedenti penali, poiché sospettati di essere membri di una gang. Si arrivò dunque a degli arresti di massa sulla base di prove circostanziali, come "mi piace" a post su Facebook che contenevano riferimenti a violenza oppure like a video musicali che parlavano di rivalità tra bande. I ragazzi arrestati furono tutti accusati di pene gravi, come associazione a delinquere e condannati ciascuno con pene differenti (Popper, B., 2014).

La vendita dei dati sensibili sui consumatori solleva preoccupazioni anche relativamente alla **sicurezza personale**.

Ad esempio, utilizzando i dettagli in possesso dei data broker sullo storico delle localizzazioni mentre si usa un'applicazione, è possibile seguire gli spostamenti di una persona e trovare il suo indirizzo di casa, esponendo la persona a rischi di stalking e molestie. Questo genere di rischio è riscontrabile anche nei prodotti people search ampiamente pubblicizzati dai data broker. Tali prodotti nascono con l'intento di facilitare la ricerca di vecchi amici, conoscenti, persone con cui si sono persi i contatti o per indagare sul passato di qualcuno e forniscono informazioni personali quali nomi, indirizzi di casa e numeri di telefono di queste persone. Tuttavia, possono essere acquistati teoricamente da chiunque, e se le aziende non controllano adeguatamente gli acquirenti, il rischio è che finiscano in possesso di "attori malintenzionati" che potrebbero utilizzare i dati con intenti dannosi. Ad esempio, espongono a ritorsioni, da parte di compagni violenti, le vittime di violenza domestica che hanno cambiato

indirizzo di casa e cercano di nascondersi (con i prodotti people search è facile rintracciare il nuovo indirizzo) oppure forze dell'ordine, procuratori e pubblici ufficiali possono sono essere esposti a minacce, stalking e vendette da parte di criminali, persone vittime di reati e non soddisfatti della giustizia ottenuta o familiari di persone coinvolti in processi penali e arrabbiati per le sentenze ricevute (FTC, 2014).

Le attività dei data broker hanno implicazioni anche sulla **sicurezza nazionale** e sul funzionamento **democratico**. Alcuni data broker come Acxiom, LexisNexis e Nielsen pubblicizzano esplicitamente i dati su attuali ed ex militari americani. Acxiom, ad esempio, possiede dati su 45,5 milioni di militari americani in servizio e pubblicizza un servizio per il marketing che consente ai propri di clienti di inviare "offerte mirate" a questa categoria. Il servizio comprende: identificazione del pubblico target a cui sarà inviata l'offerta, definizione dell'offerta del cliente (es. sconti, abbonamenti ecc.), il tipo i canali utilizzati, (e-mail, social, ecc.) e le tempistiche di lancio dell'offerta. In più Acxiom offre un servizio di "verifica e localizzazione dei militari dispiegati ma mancanti dalla base principale", indirizzato principalmente ad istituzioni finanziarie per scopi commerciali, come gestione dei crediti e dei debiti (se un militare non sta effettuando i versamenti su una carta di credito è utile localizzarlo per contattarlo e invitarlo ad effettuare i pagamenti) (Sherman, 2021).

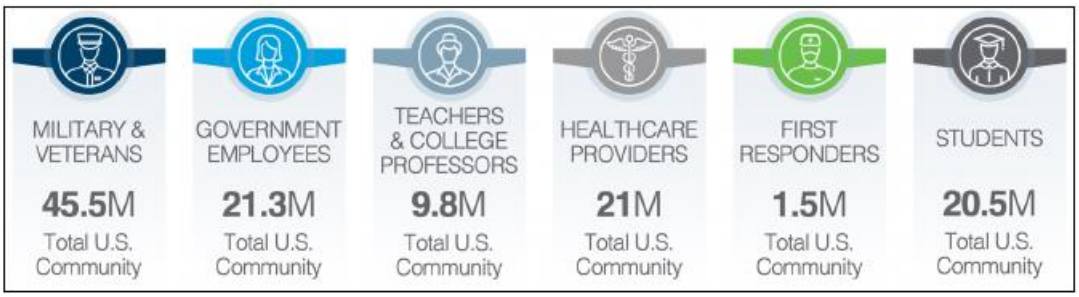

Figura 27: Pubblicità dei dati di Acxiom sul personale militare americano.
Fonte: (Sherman J., Data Brokers and Sensitive Data on U.S. Individuals, 2021)

Nielsen invece, nel 2019 ha pubblicato un report intitolato "consumatori veterani di oggi", che conteneva informazioni sulle abitudini televisive e di acquisto di militari attivi ed ex militari. Questi dati non necessariamente vengono utilizzati con intenzioni

dannose, poiché il personale militare ha caratteristiche ed abitudini uniche, molte imprese avrebbero interesse ad indirizzare offerte su misura a queste categorie, tramite l'aiuto dei data broker. Addirittura, in alcuni casi è possibile che le informazioni sui militari non siano vendute all'impresa cliente, ma gli venga concesso solo la possibilità di pubblicare annunci mirati sulla piattaforma del data broker, rivolgendo dunque tali annunci al pubblico target già individuato dal data broker. Tuttavia, in altri casi i dati sui militari sono venduti e si sa poco di queste transazioni. In America, ad esempio non esiste alcun divieto che impedisca di vendere informazioni sensibili su individui, compresi i militari, ad entità straniere, che possono utilizzare tali informazioni (cronologia di transazioni finanziarie, ricerche su Internet, localizzazione, supporto per cause e organizzazioni politiche) per creare, ad esempio, dei "profili dettagliati" di militari di alto rango con ruoli decisionali rilevanti, che una potenza straniera potrebbe utilizzare a "scopo di ricatto e coercizione" per la raccolta di informazioni, a scopo di spionaggio e cyber attacchi o addirittura, gli acquirenti combinando i dati da più data broker, possono individuare un componente della famiglia di un militare o di un funzionario del governo e ottenere la cronologia delle localizzazioni e la localizzazione in tempo reale. In quest'ultimo caso, è emblematico ciò che è successo nel 2019, quando il New York Times ha usato dati acquistati da data broker, per identificare individui appartenenti la scorta (appartenenti ai servizi segreti) del presidente Donald Trump. Nello specifico hanno ottenuto un dataset con oltre "50 miliardi di punti di localizzazione dei telefoni da oltre 12 milioni di persone" nel paese, e in pochi minuti sono riusciti **deanonimizzare**[16] i dati sulla localizzazione e tracciare la posizione del presidente Trump (utilizzando la posizione di un membro della sua scorta). Inoltre, dai dati sulla posizione, era chiaramente identificabile la casa di tale individuo che verosimilmente apparteneva ai servizi segreti, e collegando la posizione della casa agli atti pubblici, il New York Times è riuscito a scoprire il nome della persona e del suo

[16] Alcuni data broker nelle loro informative sulla privacy sostengono di avere solo dati "anonimizzati" o "deidentificati" sugli individui, ovvero rimuovono i nomi e convertono e-mail, numeri di telefono, dati sulla localizzazione, indirizzi ecc. in stringhe alfanumeriche univoche aventi tutte la stessa lunghezza. Questo processo si chiama "hashing" ed è un'operazione unidirezionale che teoricamente non può essere invertita (non si può ad esempio risalire dalla stringa alfanumerica al contenuto dei dati originario. Fonte: (Christl, W., 2017)

coniuge. Oltre ai dati sulla posizione del presidente degli Stati Uniti, hanno potuto constatare diversi punti di localizzazione legati più alla sua vita privata che ai suoi doveri pubblici. Dunque, il New York Times utilizzando i dati forniti da un data broker ha potuto localizzare un agente segreto al seguito del presidente Trump e tracciare i suoi spostamenti nell'arco di un'intera giornata, che iniziano dalla sua residenza a Mar-a-Lago Florida, per arrivare poi ad un pranzo con i leader mondiali nelle vicinanze della sua residenza e infine ad una cena privata nella sua dimora col primo ministro giapponese. (Thompson, A. et al., 2019).

La facilità disarmante con la quale è stata localizzata la posizione del presidente Trump e l'identità di uno dei membri dei servizi segreti al suo seguito, dimostra quanto vulnerabili siano i nostri dati. Ciò rappresenta una grave minaccia per la sicurezza nazionale, poiché la stessa tecnica potrebbe essere utilizzata ad esempio dalle intelligence di potenze straniere per tracciare gli spostamenti degli agenti dei servizi segreti, costruire profili di personaggi politici e membri dell'intelligence, per campagne di spionaggio nei confronti dei governi di vari paesi per ottenere insight su strategie governative, militari e geopolitiche, per sabotare operazioni militari, di intelligence o compromettere alleanze internazionali, per pianificare operazioni ostili ai governi, e infine il monitoraggio degli spostamenti di figure politiche, militari e agenti segreti può esporre questi individui ad attacchi terroristici, rapimenti o altre forme di violenza (Sherman, 2021).

Le informazioni raccolte dai data broker possono essere utilizzate inoltre per diffondere fake news e lanciare campagne di disinformazione, per fare propaganda politica e influenzare l'opinione pubblica oppure per fare profilazioni degli elettori e indirizzare messaggi mirati per influenzare il voto, minando così il libero processo democratico (Muralidhar K. et al., 2018).

Il brokeraggio di dati può anche aiutare governi stranieri a profilare gli individui per lanciare campagne di disinformazione mirate "volte a seminare il caos o a dissuadere la partecipazione degli elettori" (Sherman, 2021). In questo scenario un episodio importante riguarda l'operato nel 2016 della "Russian Internet Research Agency" (aka

IRA), azienda russa nel campo dell'informazione e della propaganda pro-Cremlino, nei confronti delle comunità afroamericane. Durante le elezioni presidenziali del 2016 l'IRA (utilizzando le informazioni dettagliate e le profilazioni sulla comunità afroamericane fornite dai data broker) utilizzò i social media per lanciare una campagna di disinformazione volta a dissuadere dalla partecipazione al voto la comunità afroamericana. I contenuti erano veicolati attraverso piattaforme quali Instagram, Facebook, Twitter e canali You Tube Oltre ai social media, l'IRA ha acquistato dei domini web come "blackmattersus.com", "black4black.info", "blacksoul.us" per veicolari i contenuti di disinformazione che si presentavano in diverse forme, come targeting pubblicitario, contenuti complottisti, manipolazione dei video e creazione di account falsi. Lo scopo di tale campagna di disinformazione era quello di favorire l'elezione di Trump, sottraendo voti a Hillary Clinton, e per fare ciò era necessario ostacolare la partecipazione al voto degli afroamericani, da sempre bacino importante di voti per i democratici (Parham, J., 2018).

Capitolo 4: Privacy e regolamentazione

Tradizionalmente la privacy, prima dell'era digitale, era intesa come "il diritto ad essere lasciati in pace" (Warren & Brandeis, 1890), evidenziando l'importanza di proteggere la propria vita privata da ingerenze esterne, in modo particolare dalla stampa. Tuttavia, la privacy non è legata solo alla protezione da ingerenze esterne, ma anche alla capacità di avere il controllo sui propri dati ed essere consapevoli di come questi dati vengono utilizzati. Infatti, secondo Westin (1967), la privacy si può considerare come "la richiesta degli individui di determinare per sé stessi quando, come e fino a che punto le informazioni che li riguardano vengono comunicate ad altri", ovvero il diritto all'autodeterminazione informativa. Nell'era digitale la sorveglianza e la raccolta massiccia dei dati sensibili operata dai data broker hanno sollevato interrogativi sulla trasparenza delle loro pratiche, con riferimento in particolare alla scarsa consapevolezza degli individui sul trattamento dei loro dati. Questi ultimi, nonostante la presenza delle informative sulla privacy, non sono pienamente consapevoli di come i data broker utilizzino i loro dati (Crain, 2018). L'asimmetria informativa tra data broker e individui può essere dovuta sia alla trasparenza limitata delle pratiche di raccolta dati che alla lunghezza e prolissità delle informative sulla privacy (gli individui, spaventati dalla complessità e dalla lunghezza, spesso accettano l'informativa senza leggere). L'asimmetria non è solo informativa ma anche tecnologica, dal momento che gli individui, a differenza dei data broker che dispongono di sofisticati tool di tracciamento, non dispongono delle competenze e degli strumenti adeguati a monitorare i propri dati e difendere opportunamente la privacy. Per aiutare gli individui ad aumentare la consapevolezza ed avere maggiore controllo sui propri dati uno strumento rilevante è stato il **GDPR** ("General Data Protection Regulation") Europeo (2016), il quale definisce la protezione dei dati personali come "un diritto fondamentale" per ogni persona "a prescindere dalla loro nazionalità o dalla loro residenza". Tale strumento normativo ha introdotto una serie di novità a favore dei soggetti "interessati" del trattamento dei dati personali, tra le quali la "**richiesta del consenso**", la "**revoca del**

consenso", il diritto alla "**portabilità dei dati**" e il diritto all' "**oblio**" (vedi par. 4.2). Tuttavia, è bene sottolineare che nonostante i numerosi sforzi fatti dai legislatori emerge una discrepanza tra le preoccupazioni manifestate dagli individui e il loro effettivo comportamento digitale; in altre parole gli individui sostengono di essere preoccupati per la condivisione dei loro dati personali, ma non fanno abbastanza per proteggerli, continuando a condividere i dati online senza essere consapevoli dei rischi a lungo termine (per avere accesso a dei servizi online o per utilizzare delle app spesso si accettano le informative o i cookie senza soffermarsi troppo su quali informazioni stiamo rilasciando). Questo fenomeno è definito in letteratura come "paradosso della privacy" (Acquisti, A. et al., 2015).

In questo capitolo si esamineranno le questioni legate alla privacy e alla regolamentazione nel mercato dei data broker. Verrà discusso un caso molto significativo di violazione della privacy e appropriazione illegittima di dati personali, ovvero il caso di Cambridge Analytica. Infine, verrà posto il focus sul **GDPR** e le relative implicazioni sul mercato dei data broker, evidenziando anche le differenze tra l'approccio normativo europeo e quello statunitense, e si discuterà delle recenti normative adottate dall'unione europea per regolamentare la circolazione dei dati, i cosiddetti **Big Data Acts**.

4.1 Privacy e violazione dei dati: il caso Cambridge Analytica

Cambridge Analytica era una società britannica nata nel 2013 con l'obiettivo di fornire consulenza politica e commerciale attraverso l'analisi dei dati (in particolare attraverso il "data mining[17]"). La società aveva sviluppato un algoritmo di "microtargeting comportamentale" che, sfruttando le "impronte digitali" lasciate dagli utenti durante la navigazione online, era in grado di prevedere le caratteristiche "emotive e

[17] È un processo che si basa sull'utilizzo di "sofisticati algoritmi matematici (tra i quali metodi di machine learning), per estrarre dati utili da ampie banche dati". Fonte: (https://www.bigdata4innovation.it/data-science/data-mining/data-mining-cose-perche-conviene-utilizzarlo-e-quali-sono-le-attivita-tipiche/)

comportamentali" di un individuo e creare dunque un profilo "psicometrico". (Della Piazza, S., 2021). L'algoritmo era stato sviluppato sulla scia degli studi dello psicologo e ricercatore Michal Kosinski, il quale sosteneva che era possibile dedurre caratteristiche sulla personalità di un individuo, basandosi esclusivamente su una certa quantità di "like" di Facebook, ad esempio con circa 68 "mi piace" è possibile prevedere orientamento politico, religioso, grado di intelligenza con accuratezza crescente col numero di "mi piace" (Kosinski, M. et al., 2013). Tale approccio ha permesso all'azienda di creare contenuti estremamente mirati e personalizzati (modulati sulla base di aspetti emotivi degli utenti), veicolati attraverso piattaforme come Facebook, in grado di influenzare le opinioni e il comportamento delle persone, ad esempio il voto degli "elettori indecisi" (Della Piazza, S., 2021).

Il caso di Cambridge Analytica è scoppiato nel 2018 quando Christopher Wylie, ex dipendente dell'azienda, ha denunciato alla stampa le pratiche poco trasparenti della società e ha rivelato in che modo erano riusciti a ottenere i dati di milioni di utenti Facebook. In un'intervista al The Guardian (2018) ha dichiarato: "Abbiamo sfruttato Facebook per raccogliere i profili di milioni di persone. E abbiamo creato modelli per sfruttare ciò che sapevamo su di loro e colpire i loro demoni interiori. Questa è stata la base su cui è stata costruita l'intera azienda" e ancora "abbiamo speso 1 milione di dollari per raccogliere milioni di profili Facebook" (Cadwalladr, C. et al., 2018). Facebook entra nella vicenda nel 2014 quando, un professore universitario di nome Aleksandr Kogan, sviluppa un'applicazione chiamata "This is your digital life", la quale si presentava come un test sulla personalità e alla fine dello svolgimento del test forniva all'utente il suo "profilo psicologico". Tuttavia, questa applicazione collezionava anche una vasta quantità di dati personali senza l'autorizzazione degli utenti. Infatti, per utilizzare l'applicazione, era necessario fare l'accesso attraverso il "Facebook Login", il quale gratuitamente permetteva agli utenti di registrarsi al sito senza creare delle nuove credenziali ma al contempo richiedeva agli utenti il rilascio di un gran numero di informazioni personali (l'applicazione, infatti, ottiene l'accesso non solo alle informazioni presenti all'interno profilo Facebook degli utenti che la utilizzano ma

anche a quelle dei profili Facebook dei loro contatti). L'applicazione fu utilizzata da circa 270 mila persone, che hanno quindi acconsentito a rilasciare le informazioni contenute nel proprio account (operazione comunque trasparente in quanto Facebook mostrava un riepilogo delle informazioni rilasciate) permettendo all'applicazione di Kogan di raccogliere e registrare informazioni, senza il consenso degli utenti (pratica che all'epoca dei fatti (2014) era ancora consentita; il GDPR sarà infatti introdotto solo nel 2018), su circa 50 milioni di profili Facebook. In seguito, questa cifra è stata rivista in eccesso da Mark Zuckenberg, il quale dichiarerà che i profili utilizzati da Cambridge Analytica per il microtargeting sono stati 87 milioni, tra cui 214 mila profili italiani (D'Alessandro, J., 2018). Il problema è sorto nel momento in cui Kogan ha venduto tali informazioni a Cambridge Analytica, in contrasto con le policy sul trattamento dei dati di Facebook, le quali vietavano espressamente, ai proprietari di app, di trasferire i dati raccolti, senza il consenso degli utenti, a società terze, pena la sospensione dell'account (Menietti, E., 2018). Secondo quanto riferito da Wylie, Facebook era perfettamente a conoscenza di tale violazione già dal 2016, tuttavia solo nel 2018, quando lo scandalo diventò di dominio pubblico, decise di sospendere dalla piattaforma Cambridge Analytica e Kogan (Della Piazza, S., 2021). Dunque, come accennato in precedenza, Cambridge Analytica aveva sviluppato un algoritmo di "microtargeting comportamentale" capace di creare dei profili molto dettagliati degli individui, sulla base anche delle loro "emozioni", col fine di indirizzare contenuti molto personalizzati a tali utenti e influenzare la loro opinione. In particolare, Cambridge Analytica, utilizzò i dati acquisiti (da 50 milioni di profili Facebook) per elaborare delle strategie di microtargeting politico su ampia scala. Infatti, nel corso delle elezioni del 2016, il comitato di Donald Trump affidò a Cambridge Analytica l'incarico di occuparsi della raccolta e analisi dei dati della campagna elettorale. L'obiettivo del comitato era di utilizzare i "profili psicografici" e il "microtargeting comportamentale" altamente puntuale di Cambridge Analytica per ottimizzare il targeting politico ed influenzare il voto degli "indecisi" attraverso annunci mirati (spesso su temi caldi quali sicurezza e immigrazione) volti a rafforzare il consenso di Trump nell'opinione pubblica (Sorte, S.,

2020). Ad oggi non è chiaro quanto e in che modo l'azienda abbia collaborato per la campagna elettorale di Trump, ma dalle indagini seguenti è emerso con certezza il coinvolgimento di Cambridge Analytica (va evidenziato inoltre che Steve Bannon, braccio destro di Trump e manager della campagna elettorale, era stato in passato vicepresidente di Cambridge Analytica) È risaputo infatti che per la propaganda pro-Trump furono ampiamente utilizzati account fake e bot con l'obiettivo di diffondere fake news e contenuti contro Hillary Clinton, spesso in tempo reale durante i dibattiti televisivi tra Trump e Clinton. Questo permetteva ad esempio di misurare la risposta dell'audience e capire quali contenuti riuscivano meglio a smuovere le emozioni degli utenti online, attività in cui Cambridge Analytica sosteneva di essere particolarmente abile (Menietti, E., 2018). Cambridge Analytica fu inoltre sospettata di aver contribuito a raccogliere dati e informazioni sugli utenti per influenzare l'opinione pubblica in merito al referendum sulla Brexit del 2016. Nello specifico, si sospettava che l'azienda avesse contatti e collaborasse con i sostenitori del "Vote Leave" (movimento a favore dell'uscita del Regno Unito dall'Unione Europea) come Nigel Farage, uno dei principali difensori della Brexit (Cadwalladr, C., 2017). L'azienda ha negato fermamente le accuse di un coinvolgimento nella campagna elettorale per la Brexit; tuttavia, rimangono forti i sospetti di una sua implicazione anche in questo caso (Menietti, E., 2018). Difatti, Wylie in una sua testimonianza del 2018 dirà ai politici britannici: "Penso che sia incredibilmente ragionevole dire che AIQ[18] ha avuto un ruolo molto significativo nella vittoria del Leave" e dirà anche "È categoricamente falso che Cambridge Analytica non abbia mai utilizzato i dati di Facebook" (Scott, M., 2018).

[18] "AggregateIQ" è una "front company" (società di facciata) canadese di Cambridge Analytica, che secondo Wylie ha contribuito insieme a Cambridge Analytica a raccogliere dati per il microtargeting degli elettori in vista del referendum sulla Brexit del 2016.

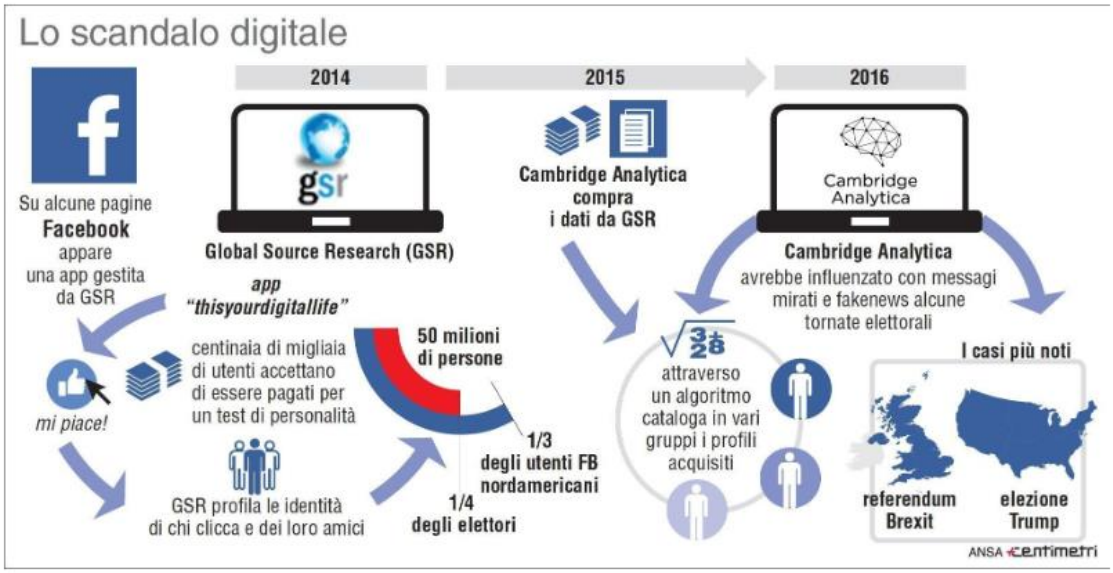

Figura 28: Il caso Cambridge Analytica-Facebook spiegato in breve.
Fonte: (https://www.cittanuova.it/dati-personali-accuse-facebook-cambridge-analytica/?ms=003&se=006)

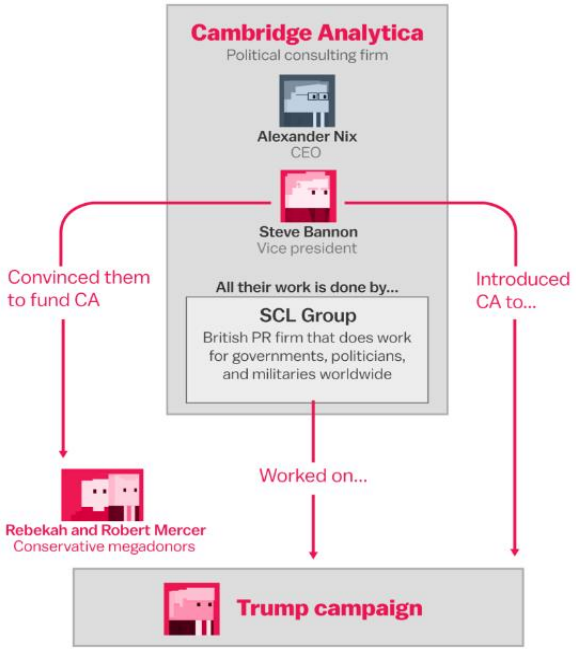

Figura 29: Legame tra campagna elettorale di Trump e Cambridge Analytica.
Fonte: (https://www.vox.com/policy-and-politics/2018/3/23/17151916/facebook-cambridge-analytica-trump-diagram)

Le conseguenze dello scandalo furono gravi e toccarono entrambe le aziende coinvolte nella vicenda. Travolta dallo scandalo, infatti, Cambridge Analytica ha dichiarato bancarotta con la conseguente cessazione delle attività e chiusura dell'azienda nel maggio del 2018 (pochi mesi dopo lo scandalo, scoppiato nel marzo 2018) (Della Piazza,

S., 2021). Facebook ha subito notevoli danni reputazionali che si sono manifestati in un forte crollo in borsa del prezzo delle azioni (alcuni giorni dopo la testimonianza di Wylie, Facebook ha chiuso col -6,7% sul NASDAQ, con il suo capitale che era sceso di circa 5 miliardi di dollari in poche ore) (Il Sole 24 Ore, 2018). Mark Zuckerberg è stato chiamato a testimoniare davanti il Congresso, il 10 e l'11 Aprile 2018, in merito al suo coinvolgimento nello scandalo. In questa occasione dirà: "è stato un mio errore, e ne sono dispiaciuto. Io ho creato Facebook, io lo mando avanti, e sono io il responsabile di ciò che accade" (Della Piazza, S., 2021) e ancora "ho chiaramente commesso un errore, avremmo dovuto fare di più" (Sky TG24, 2018) ammettendo dunque delle potenziali falle e carenze sistemiche nel meccanismo di protezione dei dati degli utenti. Facebook nel 2019 è stata multata dalla Federal Trade Commission per una cifra pari a 5 miliardi di dollari, ma nella sanzione erano previsti anche "controlli e restrizioni destinati a garantire in futuro la protezione dei dati personali degli utenti nell'ambito dei servizi e delle attività" della piattaforma (Valsania, M., 2019). In poche parole, Facebook per arrivare ad un accordo con la FTC ed evitare ulteriori testimonianze, è stato costretto a rafforzare i controlli e rendere più stringenti le proprie normative in materia di privacy (Sorte, S., 2020). Facebook non è stata sanzionata solo dalla FTC, ma anche dal Garante Per La Protezione Dei Dati Personali, il quale ha comminato alla piattaforma una sanzione di 1 milione di euro per violazione delle norme sulla protezione dei dati personali (GDPR) in merito al caso Cambridge Analytica (Garante Per La Protezione Dei Dati Personali, 2019).

Il caso ha messo alla luce le criticità legate alla diffusione e all'uso non regolamentato dei dati personali fungendo da catalizzatore per l'adozione e implementazione del GDPR in Europa (2018), il quale ha introdotto norme più stringenti sulla raccolta e condivisione dei dati, per garantire maggiore trasparenza da parte delle aziende e maggiore tutela dei diritti degli utenti (ad esempio il consenso informato degli utenti al trattamento dei dati, il diritto all'accesso e il diritto di cancellazione dei dati).

4.2 Il GDPR e le recenti normative europee

Il General Data Protection Regulation (formalmente "Regolamento Ue 2016/679") noto anche come **GDPR** è il regolamento europeo in materia di protezione dei dati personali e libera circolazione di tali dati all'interno degli Stati membri, che è stato adottato dal Parlamento Europeo il 27 aprile del 2016 ma è diventato pienamente operativo e applicabile a tutti gli Stati membri a partire dal 25 maggio 2018 (Cataleta, A. et al., 2024). Seppur in misura limitata, tale regolamento è quello che ha un impatto più significativo sui data broker (Lonardo, F., 2017). Il GDPR è composto da 99 articoli suddivisi in 11 capi, ciascuno dei quali tratta aspetti diversi della protezione dei dati personali. Il capo 1 tratta le "Disposizioni generali", come l'oggetto e l'ambito di applicazione del GDPR, le condizioni per il consenso e fornisce le definizioni chiave in materia di protezione dei dati. Anzitutto l'art. 4 del GDPR definisce "**dato personale**" come "qualsiasi informazione riguardante una persona fisica identificata o identificabile", detta "interessato"; l'"interessato" è dunque "la persona fisica che può essere identificata", anche indirettamente, attraverso "nome, numero di identificazione, dati relativi all'ubicazione, identificativi online […]". Ad esempio, dati personali possono essere: nome, cognome, codice fiscale, e-mail personale, indirizzo IP, dati sulla salute, biometrici o posizione geografica ecc., mentre dati non personali (che non rientrano quindi nel campo di applicazione del GDPR) sono: e-mail aziendale e informazioni aziendali in generale, dati statistici aggregati, numero di registrazione delle società e dati anonimizzati. Nel GDPR (art. 4) si definisce "**trattamento**, qualsiasi operazione o insieme di operazioni […] applicate a dati personali o insiemi di dati personali" come "la raccolta, la registrazione, l'organizzazione, la conservazione, la consultazione, la comunicazione mediante diffusione […]"; si definisce "**titolare del trattamento**, la persona fisica o giuridica, l'autorità pubblica, il servizio o altro organismo che, singolarmente o insieme ad altri, determina le finalità e i mezzi del trattamento di dati personali"; invece, il "**responsabile del trattamento**" è "la persona fisica o giuridica, l'autorità pubblica […] che tratta dati personali per conto del titolare

del trattamento". A titolo esemplificativo, un'azienda di e-commerce può gestire le informazioni sui propri clienti utilizzando un servizio cloud fornito in outsourcing da un'altra azienda. L'azienda di e-commerce può decidere quali e quanti dati raccogliere, la finalità e la modalità di trattamento; l'azienda è il "titolare del trattamento", mentre il fornitore del servizio cloud si limiterà solo a trattare i dati personali per conto dell'azienda, ed è il "responsabile del trattamento". Il "**destinatario**" è definito come "la persona fisica o giuridica, l'autorità pubblica […] che riceve comunicazione di dati personali […], mentre il "**terzo**" è "la persona fisica o giuridica, l'autorità pubblica […] che non sia l'interessato, il titolare del trattamento, il responsabile del trattamento […]". Infine, l'art. 4 del GDPR definisce "**consenso dell'interessato**, qualsiasi manifestazione di volontà libera, specifica, informata e inequivocabile dell'interessato, con la quale lo stesso manifesta il proprio assenso, mediante dichiarazione o azione positiva inequivocabile, che i dati personali che lo riguardano siano oggetto di trattamento".

Il GDPR ha come "**oggetto e finalità**" (art.1 GDPR) la "protezione delle persone fisiche con riguardo al trattamento dei dati personali" e "la libera circolazione di tali dati". Tale regolamento punta a proteggere "diritti e le libertà fondamentali delle persone fisiche, in particolare il diritto alla protezione dei dati personali" (art.1 GDPR) ed è stato introdotto anche per rispondere a specifiche esigenze, espresse dalla Commissione europea, di armonizzare le normative sulla privacy dei vari Stati membri e semplificare le norme relative trasferimento dei dati personali dall'Unione Europea verso paesi extra UE (Cataleta, A. et al., 2024) e, come vedremo in seguito, garantisce ai cittadini UE un maggiore controllo sull'utilizzo dei propri dati, introducendo nuovi diritti come: il **diritto all'oblio**, il **diritto alla portabilità dei dati**, il diritto di **notifica dei data breach** (le violazioni dei dati personali) (Alovisio, M., 2016). Un tema centrale nel GDPR è la libera circolazione dei dati all'interno degli Stati UE, che, secondo quanto definito dall'art. 1 par. 3, "non può essere limitata né vietata per motivi attinenti alla protezione delle persone fisiche con riguardo al trattamento dei dati personali".

L'art. 2 del GDPR regolamenta l'"**ambito di applicazione materiale**" della normativa, la quale si applica al "trattamento interamente o parzialmente automatizzato di dati

personali e al trattamento non automatizzato di dati personali", a condizione che quest'ultimi siano contenuti in un archivio strutturato o siano destinati a esservi inclusi. Sono esclusi dall'ambito di applicazione del GDPR i trattamenti di dati personali:

- "effettuati per attività che non rientrano nell'ambito di applicazione del diritto dell'Unione (ad esempio le questioni relative alla sicurezza nazionale e alla gestione delle risorse militari che rimangono di competenza esclusiva degli Stati membri)
- "effettuati da una persona fisica per l'esercizio di attività a carattere personale"
- "effettuati dalle autorità competenti a fini di prevenzione, indagine, accertamento o perseguimento di reati […]"

Lo scopo dell'art. 2 è di delineare il contesto normativo in cui si applica il GDPR e dove invece è necessario far riferimento ad altre normative. Il campo di applicazione è volutamente ampio per garantire che i dati personali dei cittadini UE siano protetti nella maggior parte delle circostanze (con le eccezioni elencate sopra).

Oltre all'ambito materiale, nel GDPR viene descritto anche l'"**ambito territoriale**" (art. 3) di applicazione della normativa. Essa si applica a qualsiasi organizzazione, avente una sede in UE, che effettua il trattamento dei dati personali (sia esso svolto dal "titolare" o del "responsabile" del trattamento), indipendentemente dal fatto che il trattamento avvenga nell'Unione o in paese extra UE. Si applica anche alle organizzazioni che hanno uno stabilimento al di fuori dell'UE, ma che trattano i dati personali di "interessati" che si trovano nell'UE.

Il GDPR prevede che il trattamento dei dati personali deve rispettare i **principi fondamentali** definiti nell'art. 5:

- i dati devono essere trattati in modo "lecito, corretto e trasparente nei confronti dell'interessato" ("**liceità, correttezza e trasparenza**");
- i dati devono essere raccolti per "finalità determinate, esplicite e legittime e […] trattati in modo che non sia incompatibile con tali finalità". La "**limitazione della finalità**" mira a garantire che le organizzazioni utilizzino i

dati solo per gli scopi esplicitamente dichiarati e che eventuali trattamenti successivi non siano in contrasto con le finalità iniziali della raccolta dei dati;

- un principio chiave della norma è la "**minimizzazione dei dati**", ovvero i dati trattati devono essere "adeguati, pertinenti e limitati a quanto necessario" rispetto alle finalità del trattamento. In altre parole, questo principio invita le organizzazioni a limitare il trattamento dei dati ai soli dati strettamente essenziali alle finalità del trattamento, riducendo così il rischio di accesso non autorizzato ai dati sensibili e migliorando di conseguenza l'efficienza nella gestione delle informazioni, la sicurezza e la trasparenza verso gli "interessati";
- i dati devono essere "esatti e, se necessario, aggiornati". Inoltre "devono essere adottate tutte le misure ragionevoli per cancellare o rettificare tempestivamente i dati inesatti rispetto alle finalità per le quali sono trattati" ("**esattezza**");
- i dati devono essere "conservati" per un arco di tempo non superiore a quello necessario al conseguimento degli scopi per i quali sono trattati ("**limitazione della conservazione**");
- deve essere garantita una sicurezza adeguata dei dati personali oggetto del trattamento, inclusa "la protezione, mediante misure tecniche e organizzative adeguate, da trattamenti non autorizzati o illeciti e dalla perdita, dalla distruzione o dal danno accidentali" ("**integrità e riservatezza**");
- una novità importante introdotta dal GDPR è il "**principio di responsabilizzazione**" o di "**accountability**" dei "titolari del trattamento", ovvero l'obbligo da parte dei "titolari" di "comprovare" che il trattamento dei dati è avvenuto secondo il rispetto delle normative del GDPR e di adottare misure tecniche e organizzative adeguate a garantire e dimostrare la conformità del trattamento ai principi del GDPR (Cataleta, A. et al., 2024);

Per garantire l'attuazione del principio di responsabilizzazione, il GDPR introduce una nuova figura chiamata "Data Protection Officer" (DPO) (art. 37), i cui compiti prevedono il supporto al titolare o al responsabile del trattamento nel garantire la conformità dei trattamenti al GDPR e il monitoraggio a tutto tondo dell'osservanza

dell'impresa o dell'ente alle disposizioni del GDPR (Cataleta, A. et al., 2024). Funge inoltre da facilitatore e comunicatore tra l'organizzazione e l'esterno, ad esempio tra organizzazione e "interessati" del trattamento o tra organizzazione e Garante della privacy. La nomina del DPO (effettuata dal titolare o dal responsabile del trattamento) è obbligatoria nel caso in cui il trattamento sia fatto da:

- "autorità pubblica o da un organismo pubblico";
- dalle aziende la cui attività principale consiste in trattamenti di dati, che "per loro natura […] richiedono il monitoraggio sistematico […] su larga scala";
- dalle aziende che effettuano trattamenti di dati su larga scala e relativi a condanne penali o reati;

Il DPO può essere nominato internamente all'organizzazione (titolare o responsabile del trattamento), oppure può essere scelto dall'esterno; in entrambi i casi il GDPR richiede che il DPO sia un "supervisore indipendente", ovvero non riceve istruzioni sulla modalità di esecuzione dei suoi compiti (Di Resta, F., 2018).

Per dar seguito al principio di "responsabilizzazione" del titolare, menzionato tra i "principi fondamentali" (art. 5) del Regolamento, il GDPR presenta anche un'altra importante novità riguardante la "responsabilità" del titolare. Quest'ultimo, infatti, eventualmente coadiuvato dal DPO, ha facoltà di redigere una "**valutazione dell'impatto**" ("Data Protection Impact Assessment") del trattamento dei dati personali, con l'obiettivo di identificare e mitigare preventivamente i rischi legati ai diritti e alle libertà degli "interessati" del trattamento. Il DPIA è una valutazione che deve essere documentata per iscritto ed è obbligatoria nei casi in cui il trattamento comporta un rischio elevato per gli interessati (l'art. 35 elenca i trattamenti considerati ad alto rischio). La redazione della DPIA, oltre ad essere segno di trasparenza e responsabilità nei confronti degli "interessati", è fondamentale per soddisfare i requisiti di "**privacy by design**" e "**privacy by default**" richiesti dal GDPR (art. 25). "Privacy by design" significa che le misure di protezione dei dati devono essere già valutate nella fase di progettazione e sviluppo dei prodotti, servizi o procedure. Il titolare deve "integrare nel trattamento le necessarie garanzie al fine di soddisfare i requisiti del […] regolamento e

tutelare i diritti degli interessati". L'approccio "privacy by default" richiede al "titolare" di individuare misure adeguate a garantire un livello di protezione elevato per gli interessati. In particolare, tali misure devono garantire che "siano trattati, per impostazione predefinita, solo i dati personali necessari per ogni specifica finalità del trattamento". Tale obbligo riguarda la "quantità dei dati personali raccolti, la portata del trattamento, il periodo di conservazione e l'accessibilità". Quindi le impostazioni di "default" della privacy devono essere configurate in automatico in modo tale che i dati personali non siano resi accessibili a un numero indefinito di persone fisiche" senza il consenso dell'utente, il quale ha in ogni caso la possibilità di modificare le impostazioni della privacy ed autorizzare un trattamento più esteso di quello strettamente necessario settato per impostazione di default.

Tornando al capo 1 ("Disposizioni generali") del GDPR, è importante evidenziare l'art. 6, il quale pone le basi giuridiche del trattamento, ovvero le condizioni sotto le quali è autorizzato legalmente il trattamento dei dati. Alcune basi giuridiche, per cui sussiste la "**liceità del trattamento"**, sono:

- il "consenso" dell'interessato;
- il "trattamento è necessario all'esecuzione di un contratto di cui l'interessato è parte […]";
- il "trattamento è necessario per un obbligo legale al quale è soggetto il titolare", per "l'esecuzione di "un compito di interesse pubblico" o è "necessario per la salvaguardia degli interessi vitali dell'interessato";

Il GDPR mette in primo piano l'importanza del "**consenso**" dell'interessato, a tal punto che viene esplicitato meglio nell'art. 7. Il titolare, dunque, non deve solo ottenere il consenso dell'interessato, ma deve poter dimostrare in modo inequivocabile che "l'interessato ha prestato il proprio consenso al trattamento dei propri dati personali" e che il consenso è stato ottenuto in modo conforme al GDPR. L'interessato deve poter prestare il proprio consenso in modo informato, per questa ragione è essenziale che il titolare presenti una richiesta di consenso facilmente "comprensibile e accessibile", utilizzando un linguaggio "semplice e chiaro". Un altro aspetto cardine è il diritto, da

parte dell'interessato, di revocare il consenso in qualsiasi momento e il processo di revoca deve essere semplice quanto quello di concessione. L'interessato deve essere informato del diritto di revoca, prima di fornire o meno il proprio consenso. In ogni caso, la revoca del consenso non "non pregiudica la liceità del trattamento" effettuato prima della revoca.

Ricordiamo inoltre che il GDPR proibisce il trattamento di alcune categorie di dati personali (art. 9), tra cui dati che "rivelino l'origine razziale o etnica, le opinioni politiche, le convinzioni religiose o filosofiche, o l'appartenenza sindacale, dati genetici, dati biometrici intesi a identificare in modo univoco una persona fisica, dati relativi alla salute o alla vita sessuale o all'orientamento sessuale della persona". Tuttavia, la normativa prevede alcune circostanze in cui il trattamento di queste categorie di dati è concesso, ad esempio:

- "l'interessato ha prestato il proprio esplicito al trattamento di tali dati […]";
- il trattamento di tali dati è necessario per "assolvere gli obblighi ed esercitare i diritti specifici del titolare del trattamento o dell'interessato in materia di diritto del lavoro e della sicurezza sociale e protezione sociale";
- il trattamento è necessario per tutelare un interesse vitale dell'interessato o di un'altra persona, qualora l'interessato sia impossibilitato a fornire il proprio consenso;
- il trattamento è necessario per motivi di "interesse pubblico rilevante", come sanità pubblica o ricerca scientifica, o per motivi legati alla "medicina del lavoro";
- il "trattamento è necessario per accertare, esercitare o difendere un diritto in sede giudiziaria";

Il capo 3 del GDPR riveste una notevole importanza in quanto introduce i "**diritti degli interessati**", i quali rivestono un ruolo centrale nella normativa. Questo capo è uno dei pilastri portanti della normativa e mira a offrire livelli elevati di tutela per i cittadini, garantendo al contempo un giusto equilibrio tra la trasparenza nel trattamento dei dati e

le necessità delle organizzazioni. A seguire, un elenco dei principali "diritti degli interessati" introdotti dal GDPR:

- "**diritto di accesso**" (art. 15): l'interessato ha il diritto di "ottenere dal titolare del trattamento la conferma che sia o meno in corso un trattamento di dati personali che lo riguardano e in tal caso, di ottenere l'accesso ai dati personali" ed altre informazioni riguardanti: "le finalità del trattamento, le categorie di dati personali in questione, i destinatari a cui i dati […] sono stati o saranno comunicati, quando possibile, il periodo di conservazione dei dati personali oppure i criteri utilizzati per determinare tale periodo". Questo diritto permette agli interessati di verificare l'accuratezza dei propri dati e intervenire eventualmente per correggerli o cancellarli;
- "**diritto di rettifica**" (art. 16): l'interessato ha il diritto di "ottenere dal titolare del trattamento la rettifica dei dati personali inesatti che lo riguardano". Il titolare è tenuto a soddisfare la richiesta "senza ingiustificato ritardo";
- "**diritto di cancellazione**" o "**diritto all'oblio**" (art. 17): gli interessati hanno facoltà di chiedere al titolare la cancellazione dei dati personali che li riguardano, se sussiste almeno una delle seguenti motivazioni: i dati personali non sono più necessari rispetto alle finalità per le quali sono stati raccolti, revoca del consenso, trattamento illecito dei dati, l'interessato si oppone al trattamento ai sensi dell'art. 21, la cancellazione a carico del titolare è imposta dalla legge. Il titolare deve adempiere alla richiesta di cancellazione "senza ingiustificato ritardo".
- "**diritto di limitazione**" (art. 18): l'interessato ha il diritto di "ottenere dal titolare del trattamento la limitazione del trattamento" quando si verifica una delle seguenti circostanze: l'interessato ne contesta l'esattezza, il trattamento è illecito e l'interessato preferisce la limitazione anziché la cancellazione oppure quando i dati non servono più al titolare ai fini del trattamento, ma sono necessari all'interessato per "l'accertamento, l'esercizio o la difesa di un diritto in sede giudiziaria";

- "**diritto alla portabilità dei dati**" (art. 20): tale diritto da all'interessato la possibilità di "ricevere in un formato strutturato, di uso comune e leggibile da dispositivo automatico" i dati personali forniti a un titolare del trattamento e "trasmettere tali dati a un altro titolare del trattamento" senza impedimenti da parte del "primo" titolare del trattamento, qualora il trattamento sia effettuato in base al consenso o tramite mezzi automatizzati Nell'esercitare il diritto alla portabilità dei dati, l'interessato ha addirittura la facoltà di ottenere la trasmissione diretta da un titolare all'altro. Il **Data Act** (formalmente "Regolamento (UE) 2023/2854"), entrato in vigore l'11 gennaio 2024 e applicabile a tutti gli Stati membri a partire dal 12 settembre 2025, estende il diritto alla portabilità dei dati a qualsiasi dato generato dall'uso di macchine e dispositivi IoT (Internet of Things). [19]
- "**diritto di opposizione**" (art. 21): da all'interessato il diritto di opporsi in qualsiasi momento, per "motivi connessi alla sua situazione particolare" al trattamento dei dati personali che lo riguardano e impone al titolare di astenersi dal trattare ulteriormente i dati personali a meno che egli non dimostri l'esistenza di "motivi legittimi cogenti" per procedere al trattamento che prevalgono sugli interessi, sui diritti e sulle libertà dell'interessato oppure per l'accertamento, l'esercizio o la difesa di un diritto in sede giudiziaria. Facendo leva su questo diritto, l'interessato può opporsi ai trattamenti per finalità di "marketing diretto", compresa la "profilazione nella misura in cui sia connessa a tale marketing diretto". L'interessato può opporsi anche al trattamento per fini

[19] Con Internet of Things si intende una rete di oggetti e dispositivi connessi a una rete Internet e dotati di sensori che permettono loro di trasmettere e ricevere dati, da e verso altre cose e sistemi. Quindi tra l'IoT rientra qualsiasi tipologia di oggetti "intelligenti" tra loro interconnessi e capaci quindi di scambiare tra loro le informazioni possedute, raccolte ed elaborate. Tra i campi di applicazione dell'IoT vi sono: le auto intelligenti (ad esempio le vetture 5G, le reti energetiche intelligenti (ad esempio i lampioni dotati di sensori che consentono di rilevare i problemi e avvisare i tecnici per le riparazioni), le "smart home" (sistemi di automazione domestica per gestire da remoto l'illuminazione o gli elettrodomestici), la "smart agriculture" (sensori per rilevare i parametri climatici e le condizioni del suolo) e la "telemedicina". Fonte :(https://blog.osservatori.net/it_it/cos-e-internet-of-things)

di ricerca “scientifica e storica” o a fini statistici, salvo se tale tipo di trattamento è “necessario per l'esecuzione di un compito di interesse pubblico”;

Ricordiamo brevemente gli art. 13 e 14 del GDPR che normano le “**informative sulla privacy**”, cioè le informazioni che il titolare del trattamento deve fornire agli interessati nel momento in cui i dati personali sono stati ottenuti. Tra le informazioni da fornire figurano: l'identità e i dati di contatto del titolare del trattamento o del suo rappresentante, dati di contatto del DPO se nominato, finalità del trattamento e base giuridica, legittimi interessi perseguiti dal titolare, eventuali destinatari dei dati personali, periodo di conservazione dei dati, diritti dell’interessato tra cui il diritto di revoca del consenso e il diritto di fare reclamo all’autorità di controllo (Garante della privacy) ecc.

Tra gli altri articoli del GDPR ricordiamo:

- **L’articolo 30** che obbliga i titolari a tenere un “**registro dei trattamenti**”, contenente i dettagli su tutte le attività di trattamento svolte dal titolare. Alcune delle informazioni da conservare nel registro sono: contatti del titolare, finalità del trattamento, destinatari dei dati, descrizione delle categorie di interessati e dati personali, eventuali trasferimenti di dati verso un paese terzo o un’organizzazione internazionale compresa l'identificazione del paese terzo o dell'organizzazione internazionale, termini previsti per la cancellazione dei dati e descrizione generale delle misure di sicurezza tecniche e organizzative;
- **L’articolo 32** afferma che il “titolare del trattamento e il responsabile del trattamento” devono mettere in atto “misure tecniche e organizzative” idonee a “garantire un livello di sicurezza adeguato al rischio”, ad esempio di accesso non autorizzato, distruzione accidentale o illecita. Tali misure devono essere tarate in base alla rischiosità del trattamento dei dati specifico e possono comprendere: “pseudonimizzazione e cifratura dei dati personali, assicurare su base permanente la riservatezza, l'integrità, la disponibilità e la resilienza dei sistemi e dei servizi di trattamento, ripristinare tempestivamente la disponibilità e l'accesso dei dati personali in caso di incidente fisico o tecnico, procedure per

testare, verificare e valutare regolarmente l'efficacia delle misure tecniche e organizzative;

- **Gli articoli 33 e 34:** l'art. 33 ("notifica di una violazione dei dati personali all'autorità di controllo") stabilisce che il titolare, in caso di violazione dei dati personali, ha l'obbligo di notificare la violazione al Garante della privacy, "senza ingiustificato ritardo e ove possibile entro 72 ore dal momento in cui ne è venuto a conoscenza". Se la violazione è correlata a trattamenti svolti dal "responsabile", quest'ultimo è tenuto a informare della violazione il titolare "senza ingiustificato ritardo". Il GDPR (art. 4 par. 12) definisce "**violazione dei dati personali**" qualsiasi evento che "comporta accidentalmente o in modo illecito la distruzione, la perdita, la modifica, la divulgazione non autorizzata o l'accesso ai dati personali trasmessi, conservati o trattati".

L'art. 34 invece impone al titolare del trattamento di notificare la violazione dei dati all'interessato, quando tale violazione è "suscettibile di presentare un rischio elevato per i diritti e le libertà" dell'interessato, senza alcun "ingiustificato ritardo". La violazione va comunicata all'interessato utilizzando un linguaggio "semplice e chiaro" e la comunicazione deve contenere almeno le informazioni di cui all'art. 33 par. 3, ad esempio: la "natura della violazione" e se possibile una stima della categoria e del numero degli interessati e dei dati in questione dati di contatto del DPO, le probabili conseguenze per l'interessato della violazione, le misure adottate o da adottare per attenuare gli effetti negativi della violazione. Esistono delle situazioni in cui la comunicazione all'interessato non è obbligatoria, ovvero:

a) quando il titolare del trattamento ha messo in atto "tecniche e organizzative adeguate di protezione e tali misure erano state applicate ai dati personali oggetto della violazione". Tali misure includono, ad esempio, la "cifratura", che rende i dati personali incomprensibili a chiunque non abbia l'autorizzazione all'accesso;

b) quando il titolare del trattamento, in seguito alla violazione, ha adottato misure atte a scongiurare il sopraggiungere di un rischio elevato per i diritti e le libertà degli interessati;
c) se la notifica richiede "sforzi sproporzionati". In tal caso può essere fatta una comunicazione pubblica o una misura simile con la quale gli interessati vengono informati con eguale efficacia;

Il mancato rispetto delle norme del GDPR può portare a sanzioni amministrative pecuniarie (art. 83 e 84), che vengono poi comminate dal Garante della privacy. Tali sanzioni devono essere "effettive, proporzionate e dissuasive". Vengono classificate in base a diversi fattori, tra i quali: "natura, gravità e durata della violazione", se la violazione è dolosa o colposa, l'adeguatezza delle misure adottate dal titolare ecc. Ad esempio, per le cosiddette "violazioni di minore gravità" (come gli "obblighi del titolare del trattamento" a norma degli art. 8,11, da 25 a 39, 42,43 oppure gli "obblighi dell'organismo di certificazione" a norma degli art. 42 e 43) sono previste sanzioni fino a 10 milioni di euro, o per le imprese con fatturato superiore, la multa è del 2% del fatturato dell'esercizio precedente. Oppure per le violazioni più gravi, come le violazioni dei "diritti degli interessati" (art. da 12 a 22) le sanzioni ammontano fino a 20 milioni di euro, o se il fatturato è superiore, fino al 4% del fatturato totale dell'esercizio precedente. Il GDPR non prevede direttamente sanzioni penali dirette, ma lascia agli Stati membri la possibilità di introdurre sanzioni penali a livello nazionale, per le violazioni del GDPR.

Tra le altre normative europee in materia di privacy vi è il **Data Act** (formalmente "Regolamento (UE) 2023/2854), entrato in vigore l'11 gennaio 2024 ma operativo e applicabile solamente a partire dal 12 settembre 2025. Il Data Act è la norma più recente che va a completare la "strategia europea sui dati", comprensiva già dei seguenti regolamenti: **Digital Services Act**, **Digital Markets Act** e **Data Governance Act**. Il Data Act stabilisce "norme armonizzate per l'utilizzo e l'accesso equo ai dati (anche non personali)" e mira a creare un ecosistema di dati innovativo, efficiente e trasparente per

cittadini, imprese e pubbliche amministrazioni, oltre a mirare a una gestione dei dati più uniforme in Europa (Cataleta et al., 2024). L'introduzione di tale regolamento è da vedersi nell'ottica di una volontà, da parte dell'Unione Europea, di rafforzare la **sovranità digitale** e ridurre la propria dipendenza dai giganti tecnologici extraeuropei, come quelli statunitensi. Infatti, la "**strategia europea sui dati**" prevede anche la creazione di uno "**spazio comune europeo dei dati**", ovvero un'infrastruttura unica per la condivisione dei dati tra diversi settori e tra organizzazioni pubbliche e private, all'interno dei paesi dell'Unione Europea. Lo "spazio comune europeo dei dati" è un'iniziativa che mira a facilitare la condivisione e l'accesso ai dati per pubblico e privato all'interno dell'UE tutelando al contempo la privacy e i diritti dei cittadini. Ad esempio, una recente proposta dell'Unione Europea, mira a introdurre uno "spazio comune europeo sui dati sanitari", detto "EHDS (European Health Data Space). Queste iniziative suggeriscono che i dati non vengono più visti dall'Europa come una risorsa esclusivamente privata ma come un bene strategico da utilizzare per promuovere l'innovazione tecnologica, la digitalizzazione e una cultura "data driven" in tutta Europa (Cataleta et al., 2024).

Il Data Act si applica a: fornitori di "prodotti connessi" (prodotti Iot) immessi sul mercato UE e fornitori di "servizi correlati" indipendentemente da dove questi hanno sede, ai titolari di dati che mettono i dati a disposizione dei "destinatari" nell'UE, ai "destinatari" a cui dati sono stati messi a disposizione e ad "enti pubblici, Banca centrale europea e organismi dell'Unione che richiedono i dati ai titolari, nel caso essi siano "necessari", a fronte di "necessità eccezionali", per l'esecuzione di un compito nell'interesse pubblico e ai titolari che forniscono tali dati (art. 1 par. 3).

Tra le novità introdotte dal Data Act si ricordano:

- il **diritto alla portabilità "rafforzato"**: come accennato in precedenza, la portabilità dei dati viene estesa a qualsiasi dato generato da macchine e dispositivi IoT; in altre parole, grazie al Data Act basta solo chiedere ad un'azienda di trasferire i dati ad un'altra azienda che offre un servizio analogo. Ad esempio, col GDPR, non era possibile applicare il diritto alla portabilità dei

dati a trattamenti necessari per l'esecuzione di un compito pubblico ed il diritto era applicabile solo ai dati personali, mentre col Data Act il diritto alla portabilità è esteso a qualsiasi tipo di dato generato da dispositivi connessi (fermo restando le disposizioni dell'art. 20 del GDPR);

- l'obbligo di "**rendere accessibili all'utente**[20] **i dati del prodotto e dei servizi correlati**": i dati contenuti nei devices connessi devono essere, per impostazione predefinita, accessibili all'utente in modo "facile, sicuro, gratuito, in un formato completo, strutturato, di uso comune e leggibile da dispositivo automatico e, ove pertinente e tecnicamente possibile, in modo diretto" (art. 3 par. 1);
- gli utenti hanno diritto di accesso e utilizzo ai "**dati generati dall'uso di prodotti connessi o servizi correlati**". Se gli utenti non riescono ad accedere direttamente ai dati, i titolari del trattamento mettono a disposizione dell'utente i dati e i metadati[21] necessari per la loro interpretazione, "senza indebito ritardo" (art. 4);
- "**diritto dell'utente di condividere i dati con i terzi**": su richiesta di un utente, il titolare dei dati mette a disposizione di terzi i dati generati dall'uso di un prodotto o di un servizio correlato, nonché i pertinenti metadati "senza indebito ritardo" e in modo "facile, sicuro, a titolo gratuito per l'utente, in un formato completo, strutturato, di uso comune e leggibile da dispositivo automatico" (art. 5);
- I titolari del trattamento, quando devono mettere i dati a disposizione di un destinatario, sono tenuti a farlo a condizioni "eque, ragionevoli, non discriminatorie e in modo trasparente" (art. 8);

[20] È la persona fisica o giuridica che possiede un prodotto connesso e ha temporaneamente i diritti di utilizzo di tale "prodotto connesso" e riceve un "servizio correlato". Fonte:(art. 2 par. 12 Data Act)
[21] È una descrizione strutturata del "contenuto o dell'uso dei dati che agevola la ricerca o l'utilizzo di tali dati". Ad esempio, data di creazione, formato, dimensione, proprietà, origine dei dati ecc. Fonte:(art. 2 par. 2 Data Act)

- Il titolare del trattamento è obbligato a “mettere a disposizione i dati e i metadati pertinenti, a: enti pubblici, Commissione europea, Banca centrale europea e organismi UE”, quando questi dimostrano “una necessità eccezionale (stabilite dall’art. 15)” di utilizzare tali dati (art. 14);
- Gli articoli da 23 a 30 normano il passaggio dei clienti da un fornitore di servizi di trattamento (fornitori di piattaforme cloud ad esempio) ad un altro. Lo scopo di questi articoli è di ridurre gli ostacoli che riguardano i cambi di fornitori e di facilitare al contempo la portabilità e l’interoperabilità tra le piattaforme (come quelle di cloud). Esiste infatti un effetto “lock-in” per cui gli utenti hanno dei vincoli tecnologici che gli impediscono di cambiare fornitore, senza perdere le proprie informazioni. Il Data Act mira a ridurre questo effetto;
- Gli articoli da 33 a 36 stabiliscono i “requisiti essenziali” per facilitare l’“interoperabilità dei dati” e regolano i “contratti intelligenti” (utilizzati nella blockchain), ovvero programmi informatici che eseguono automaticamente le condizioni stabilite da un contratto, senza necessità di un’intermediaria, non appena vengono soddisfatti i termini stabiliti tra le parti;
- L’art. 37 stabilisce che ciascuno stato membro deve designare “una o più autorità competenti” o in alternativa fare affidamento sulle “autorità esistenti”, che devono occuparsi dell’“applicazione ed esecuzione” del regolamento. Per la commissione Europea, la Banca Centrale e altri organismi UE l’autorità incaricata è il “Garante europeo della protezione dei dati”;

In alcuni casi il Data Act integra e completa il GDPR (come nel caso del diritto alla portabilità “rafforzato”), in altri casi lo limita. Tuttavia, in caso di conflitti tra GDPR e Data Act, prevalgono le disposizioni del GDPR.

In conclusione, mentre il GDPR si concentra prevalentemente sulla protezione della privacy, il Data Act offre un quadro normativo che cerca di promuovere l’innovazione e la crescita, cercando di creare un’economia europea dei dati che sia sicura, accessibile,

vivace, competitiva e trasparente e mantenendo di pari passo l'innovazione e la sovranità digitale con la tutela dei diritti dei cittadini.

La strategia europea sui dati si compone anche del cosiddetto **Digital Markets Act** (DMA), che insieme al **Digital Services Act** (DSA), compone il **Digital Services Package**.

Il **Digital Markets Act** è stato approvato dall'Unione Europea il 5 luglio 2022, ed è diventato operativo nel 2023, con l'obiettivo di disciplinare il comportamento delle grandi piattaforme online, definite nel regolamento "**gatekeeper**", ovvero social network, browser, servizi di messagistica e piattaforme online in generale che, per dimensione e rilevanza, esercitano un'influenza dominante e detengono da sole la quota maggiore del mercato. In particolare, il Digital Markets Act mira a contrastare gli abusi di posizione dominante da parte delle grandi piattaforme online, **prima** che si verifichi l'abuso. È uno strumento normativo basato su un approccio ex ante; infatti, vengono introdotti una serie di regole e disposizioni che cercano il contrastare le pratiche sleali che possono essere messe in atto dai gatekeeper. Si differenzia dalle normative antitrust che agiscono ex post, ovvero la sanzione è comminata dopo che la violazione anticoncorrenziale è stata messa in atto. Il Digital Markets Act presenta sia divieti, contenuti in una sorta di blacklist delle pratiche vietate per evitare comportanti anticoncorrenziali sia nuovi obblighi per le aziende. Ad esempio, i gatekeeper non potranno (Cataleta et al., 2022):

- sfruttare la propria posizione dominante sul mercato per monopolizzare nuovi mercati, imponendo ad esempio commissioni elevate;
- favorire eccessivamente i propri prodotti sulle piattaforme a scapito di quelli proposti da altre aziende;
- rifiutarsi di concedere la portabilità dei dati o il riuso all'utente, per disincentivare quest'ultimo ad abbandonare la piattaforma;
- rifiutarsi di concedere a terze parti l'accesso ai dati dell'utente, se quest'ultimo ha concesso l'autorizzazione;

- ostacolare l'interoperabilità tra servizi per disincentivare l'utente ad abbandonare la piattaforma;
- imporre limitazioni ingiustificate sugli utenti nell'utilizzo di servizi o applicazioni di terze parti;

Figura 30: Principali divieti del DMA per i gatekeeper.
Fonte:(https://www.consilium.europa.eu/media/55547/sn01036-re02-it22.jpg).

I gatekeeper, tra i loro obblighi, hanno anche le seguenti disposizioni: devono concedere agli utenti la possibilità di disinstallare qualsiasi applicazione software, installata in automatico al momento dell'iscrizione alla piattaforma, devono evitare di garantire spazi pubblicitari favorevoli ai prodotti che appartengono all'impresa a discapito dei

prodotti altrui, e fornire l'interoperabilità tra i loro servizi e quelli dei concorrenti in modo tale da favorire la concorrenza nel settore.

Gli obiettivi del DMA sono quelli di eliminare le barriere all'ingresso e combattere gli abusi delle grandi piattaforme digitali, per arrivare ad un mercato digitale equo e competitivo e per stimolare la crescita, l'innovazione e l'espansione di tali mercati, permettendo ad esempio alle piccole imprese e start-up innovative di competere con i big del mercato (Condemi, J., 2022).

Il **Digital Services Act** (DSA) è stato approvato dall'Unione europea il 19 ottobre 2022, poco tempo dopo il Digital Markets Act. Il Regolamento è diventato esecutivo nel 2023 inizialmente per 17 Very Large Online Platforms (VLOPs) e 2 Very Large Online Search Engines (VLOSEs), individuate dalla Commissione Europea. Tra queste vi figurano: Facebook, Instagram, Twitter, Amazon Store, Google Play, Booking.com, You Tube, Tik Tok, Zalando, Bing, Google Search ecc. Tuttavia, il DSA, dal 17 febbraio 2024, è diventato applicabile e valido per tutte le piattaforme digitali. Tale Regolamento prevede un quadro normativo per regolare la trasparenza, la responsabilità delle piattaforme online e la moderazione dei contenuti, per proteggere i diritti dei consumatori e per contrastare la diffusione di contenuti illegali e la disinformazione. Si applica a social network, mercati online, app store, servizi di cloud, fornitori Internet ecc. ed è basato sul pattern: "ciò che è illegale offline dovrebbe essere illegale anche online" (Consiglio dell'Unione europea, 2021). Il Regolamento suddivide le piattaforme in quattro categorie: servizi di intermediazione (provider Internet), hosting (es. piattaforme cloud), piattaforme online (social media), "very large online platform" (VLOPs) insieme ai "very large search engines" (VLOSEs), con ciascuna categoria che deve assolvere a degli obblighi specifici, proporzionati alla tipologia di piattaforma e alla sua rischiosità per gli utenti. Tra gli obblighi più importanti ricordiamo:

- presentare in modo chiaro e trasparente le condizioni di servizio;

- la segnalazione alle autorità competenti e la rimozione di contenuti illegali in modo rapido e tempestivo. Le piattaforme devono avere inoltre procedure chiare e ben definite che permettano agli utenti di segnalare i contenuti illeciti;
- fornire informazioni esplicite sulla moderazione dei contenuti e sui propri algoritmi, soprattutto quelli che influenzano i suggerimenti e le pubblicità rivolte agli utenti;
- divieto di impiegare tecniche di "targeting" che possono "rivelano o inferiscono i dati personali dei minori o delle persone vulnerabili" ai fini di proporre pubblicità mirata ai minori o alle persone vulnerabili (art. 24);
- collaborare con le autorità nazionali in caso di indagini, se richiesto;
- dare la possibilità agli utenti di contestare la rimozione dei contenuti, se essi ritengono sia ingiusta;

Le Very Large Online Platform (VLOPs), cioè le piattaforme con più di 45 milioni di utenti attivi al mese, hanno obblighi più rigorosi perché presentano rischi elevati per i cittadini. Tali obblighi riguardano:

- valutazione periodica e prevenzione dei rischi sistemici;
- adesione a codici di condotta specifici;
- obbligo di sottoporsi ad audit indipendenti;
- fornire informazioni chiare ed esplicite circa il funzionamento dei propri algoritmi e fornire eventualmente tali dati all'autorità, se questi sono necessari a valutazioni indipendenti sul rispetto del DSA;
- dare agli utenti la possibilità di bloccare i suggerimenti (di pubblicità) basate sulla profilazione;

Le segnalazioni delle violazioni vanno fatte direttamente alla piattaforma, che ha il dovere di prenderla in carico e decidere rapidamente le azioni correttive da intraprendere (ad esempio rimuovere o vietare per sempre il contenuto). Il DSA istituisce inoltre la figura del "Trusted Flagger (art. 61)" (segnalatore affidabile), che può essere una singola persona o un'organizzazione riconosciuta ufficialmente da uno Stato UE, che ha il compito di identificare e segnalare rapidamente alle piattaforme i contenuti illeciti. L'art.

38 introduce il ruolo del **Digital Services Coordinator**, ovvero un'autorità nazionale indipendente con compiti di vigilanza e ispezione sull'applicazione del regolamento. Il suolo ruolo prevede anche l'imposizione di sanzioni alle piattaforme, il coordinamento nazionale sulle norme e funge da punto di contatto con i coordinatori di altri Stati europei per questioni che riguardano piattaforme transfrontaliere. I Digital Services Coordinator di ciascuno Stato europeo formano il "Comitato europeo per i servizi digitali" (EBDS) il quale ha funzione di supporto al coordinamento tra i paesi UE e di controllo sulle grandi piattaforme (VPOs) In conclusione, il DSA mira a creare una data economy sicura e trasparente, che coniughi la tutela dei diritti dei cittadini e al contempo l'innovazione e la competitività (Condemi, J., 2022).

La "strategia europea sui dati" si compone, infine, del **Data Governance Act** (DGA), approvato nel maggio del 2022, e diventato pienamente operativo e applicabile a partire da settembre 2023. Tale regolamento è il pilastro fondamentale per il corretto funzionamento dello "**Spazio comune europeo dei dati**[22]" in diversi settori quali "sanità, mobilità, servizi finanziari, energia e clima, spazi europei di dati per la pubblica amministrazione" ecc. Tale iniziativa è stata fortemente voluta dall'Unione europea e attualmente è ancora in fase di sviluppo. Il DGA determina le linee guida e le condizioni che l'Europa ha adottato in materia di condivisione dei dati (personali e non) tra soggetti pubblici e privati (aziende, enti pubblici, cittadini ecc.), individuando tre capisaldi: il "riutilizzo" (cioè, uso per scopi diversi da quelli iniziali) dei dati pubblici, i fornitori di servizi di intermediazione di dati e l'altruismo dei dati. Il DGA mira ad estendere il perimetro di aziende e organizzazioni che possono avere accesso e utilizzare i dati appartenenti agli enti pubblici. Il "riutilizzo" di tali dati può essere importante per le piccole imprese e le start-up che, sfruttando i dati, possono migliorare le proprie attività

[22] La "European Data Spaces" è un'iniziativa dell'Unione Europea presentata insieme alla "European Data Strategy" nel 2020. L'idea è la creazione di infrastrutture comuni per la condivisione dei dati, organizzati per settori specifici come sanità, finanza, mobilità, ambiente, agricoltura, energia e pubblica amministrazione. La "Strategia Europea sui Dati" mira a creare un mercato unico europeo dei dati entro il 2030, con lo scopo di favorire l'interoperabilità e la condivisione dei dati tra diverse organizzazioni e aziende all'interno dell'UE. L'UE vuole che i dati generati all'interno dell'Unione siano valorizzati e possano essere utilizzati per sviluppare nuovi prodotti, servizi e soluzioni innovative per migliorare e rafforzare la competitività europea rispetto ai big del settore che dominano a livello globale. Fonte :(https://digital-strategy.ec.europa.eu/en/policies/data-spaces)

e potenziare l'innovazione. La normativa consente agli enti pubblici di applicare delle tariffe per l'accesso e il riutilizzo dei dati e possono concedere esenzioni, ad esempio, per le piccole e medie imprese. Inoltre, grazie alla normativa, le aziende possono accedere a una vasta gamma di dati e hanno maggiore controllo sui propri dati, compresa la facoltà di decidere come saranno utilizzati da terzi. Il DGA norma anche le procedure con cui possono essere offerti i servizi di intermediazione dei dati, ovvero servizi che facilitano condivisione dei dati tra le diverse parti coinvolte all'interno di uno spazio condiviso. Quest'ultimi dovranno comunicare all'Autorità competente la loro attività, ai sensi dell'art. 12. Il fornitore di servizi riceverà, ad esempio, dall'autorità competente un logo ufficiale che utilizzerà nelle comunicazioni, e sarà chiaramente identificabile dall'utente come "Fornitore di servizi di intermediazione dei dati riconosciuti nell'Unione". Il terzo pilastro del Regolamento è l'"altruismo dei dati", cioè la messa a disposizione dei dati su base volontaria, da parte degli interessati stessi che hanno fornito il consenso, per iniziative di interesse pubblico e di bene comune, tra i quali la lotta al cambiamento climatico, per la divulgazione scientifica e statistica, per migliorare la qualità dei servizi pubblici o delle politiche pubbliche ecc. Gli articoli dal 16 al 25 discutono le condizioni per agevolare l'altruismo dei dati da parte degli Stati membri. Ad esempio, sono state istituite le "Organizzazioni per l'altruismo dei dati", ovvero soggetti che fungono da intermediari tra i titolari dei dati[23] e i soggetti che li utilizzeranno. Viene inoltre introdotto lo "sportello unico", ovvero un sistema centralizzato in cui soggetti pubblici o privati possono inoltrare le richieste per il riutilizzo dei dati pubblici e chiedere informazioni. Lo sportello riduce la complessità burocratica, ad esempio di dover interagire con più enti pubblici, e fornisce uno spazio unico in cui richiedere le informazioni, inoltrare le richieste di utilizzo ed in generale inviare e ricevere comunicazioni verso gli enti pubblici in modo fluido. Lo scopo della normativa è di agevolare la condivisione e l'accesso ai dati, a soggetti pubblici e privati,

[23] "La persona giuridica o l'interessato che, conformemente al diritto dell'Unione o nazionale applicabile, ha il diritto di concedere l'accesso a determinati dati personali o non personali sotto il proprio controllo o di condividerli". Fonte: (art. 2 par.5 Data Governance Act)

promuovendo un accesso equo e trasparente e garantendo comunque la protezione dei dati personali (Cataleta, A., 2023).

4.3 Differenze legislative tra Europa e Stati Uniti

Negli Stati Uniti e nell'Unione europea la privacy è trattata in modo diverso e queste differenze possono influire e incidere in modo distinto sui data broker. In generale condividono alcuni principi comuni sulla privacy, noti come "Fair Information Practice Principles" (FIPPs), che costituiscono un "framework" ampiamente accettato nel mondo. Questi principi riguardano la raccolta e l'uso dei dati, e includono: la "minimizzazione" dei dati e l'utilizzo dei dati solo per determinati casi. Tuttavia, sono generici e lasciano grande spazio all'interpretazione. L'approccio alla privacy negli Stati Uniti si basa sul concetto che il trattamento dei dati è consentito, a meno che non sia diversamente indicato da una legge specifica di un determinato Stato. In Europa invece tutti i trattamenti dei dati personali devono essere conformi ai regolamenti e sono soggetti ad obblighi di "trasparenza e legittimità". Inoltre, la privacy è riconosciuta in Europa come un "diritto fondamentale".

Negli Stati Uniti non esiste una legge federale che regoli la collezione e la vendita dei dati di entità commerciali, come i data broker. Invece vi sono una serie di leggi settoriali che regolano il trattamento dei dati per alcuni soggetti, per determinate situazioni o per alcuni settori come: i report di credito, informazioni sulla salute, dati bancari, informazioni sulle proprietà immobiliari ecc. Il risultato di ciò è un quadro normativo in materia di privacy che è "disordinato e pieno di lacune". Tra queste normative abbiamo il **Fair Credit Reporting Act** (FCRA) (1970), che stabilisce linee guida per i data broker che vendono dati sui consumatori, che sono poi utilizzati dalle banche o dalle imprese per decidere sul credito (report di credito) o sull'assegnazione di un impiego. Nello specifico tale Regolamento vieta alle aziende di fornire report sui consumatori e report sul credito a terzi, a meno che non abbiano uno scopo che rientra tra quelli consentiti dalla legge. Inoltre, le aziende hanno l'obbligo di avvisare i consumatori se un report

viene utilizzato da una banca o un'assicurazione per intraprendere un'azione negativa nei loro confronti (negare un prestito o un'assicurazione) e devono adottare misure adeguata alla tutela della privacy dei propri clienti e dei consumatori. Tuttavia, il Regolamento ha un campo di applicazione limitato; non si applica infatti ai data broker che vendono dati per scopi diversi da quelli coperti dalla legge (ad esempio per il marketing). Poi abbiamo una legge che non influisce direttamente sui data broker, ma sui tipi di dati richiesti e usati per il decision making dai clienti dei data broker. Questa norma è **l'Equal Credit Opportunity Act** (ECOA) (1974) che mira a impedire agli istituti di credito di negare un prestito ad un richiedente, su "basi proibite", come la razza o l'età del mutuatario. Per questo i data broker, rimangono vigili sui tipi di report di credito che vendono. In modo molto simile, vi è il **Fair Housing Act** (FHA) (1968), il quale protegge gli individui da discriminazioni quando acquistano o affittano una casa. Tale legge vieta espressamente di "creare, stampare o pubblicare, o far creare, stampare o pubblicare qualsiasi avviso, dichiarazione o pubblicità, in merito alla vendita o all'affitto di un'abitazione che indichi qualsiasi preferenza, limitazione o discriminazione basata su razza, colore, religione, sesso, handicap, stato familiare o origine nazionale".
La Federal Trade Commission è l'autorità che si avvicina di più ad un'autorità di protezione dei dati (in Europa vi è il Garante della Privacy europeo che compiti specifici di monitoraggio dell'esecuzione dei regolamenti ed eventualmente sanzionatorio). Può citare in giudizio imprese per pratiche anticoncorrenziali o per le loro pratiche scorrette sulla privacy; tuttavia, i suoi poteri limitati non gli permettono di controllare e monitorare ex ante le pratiche scorrette, ma solo le pratiche che causano un danno diretto ai consumatori. Sebbene ci siano stati molti sostenitori di una normativa federale sulla privacy e sul commercio dei dati personali, tali sforzi non si sono tramutati in un qualcosa di concreto. La proposta di Obama, ad esempio, del "Consumer Privacy Bill of Rights" è stata fermata dai rappresentanti dei settori ed è molto probabile, vista l'ascesa delle lobby del settore tecnologico, che una legislazione federale sarà improbabile anche in futuro. In Europa al contrario, si dispone di un quadro normativo sulla privacy che si applica a tutti gli Stati membri dell'UE e qualsiasi soggetto pubblico

o privato. La privacy è infatti riconosciuta come un diritto fondamentale, secondo quanto stabilito dal GDPR, che stabilisce obblighi rigorosi per il trattamento dei dati personali. Rimane il fatto che non esiste una normativa che regola esplicitamente i data broker e come accennato, in precedenza, il GDPR ha problemi di enforcement nel contesto dei data broker e quindi le autorità pubbliche hanno difficoltà ad applicarlo (Rieke, A. et al. 2016).

Tuttavia, mentre in Europa c'è la volontà e un contesto normativo che mira a limitare l'operato dei data broker e tutelare la privacy dei cittadini, negli Stati Uniti i data broker operano senza alcuna limitazione. Sono state fatte recentemente diverse proposte interessanti di legge, poi bocciate dal Congresso. Tra queste ricordiamo: il Data Broker List Act (2019) che avrebbe imposto ai data broker di iscriversi ad un registro nazionale, supervisionato dalla FTC, e di mantenere un programma di sicurezza completo per proteggere i dati dei consumatori, il Data Broker Accountability and Transparency Act (2020) che avrebbe imposto ai data broker di garantire agli utenti opzioni di opt-out, ovvero di opporsi al trattamento e richiedere eventualmente la cancellazione dei dati, l' Information Transparency & Personal Data Control Act (2021) che avrebbe richiesto ai data broker di ottenere necessariamente il consenso degli utenti al trattamento dei dati sensibili e di sottoporsi ad audit annuali sul livello degli standard di privacy e infine il Data Accountability and Trust Act che "avrebbe stabilito standard di sicurezza e richiesto verifiche post violazione dei dati nei confronti dei data broker oltre a vietare la raccolta di informazioni con falsi pretesti" (Reviglio, U., 2022).

Una delle leggi più importanti è la legge della California, il California Consumer Privacy Act (2018) che ha introdotto nuovi diritti per i "consumers" (interessati del trattamento) residenti in California. In modo simile al GDPR, ha introdotto diritti per i consumers, riguardanti l'accesso ai dati, la portabilità, diritto alla cancellazione, diritto a essere informato sulle categorie di dati personali trattati e il diritto di impedire la vendita dei dati personali a terzi (Belfi, M., 2020).

4.4 Impatto delle normative europee sui DB

Ruschemeier (2023) nel suo articolo "*Data Brokers and European Digital Legislation*" analizza i data broker e i loro modelli di business e valuta l'effetto delle nuove normative europee (DA, DMA, DSA, DGA) su di loro. Il GDPR, sebbene sia una delle normative più avanzate e complete in termini di protezione della privacy non affronta direttamente l'argomento data broker e incontra difficoltà nell'applicare le proprie disposizioni alle specificità del business dei data broker, specialmente quelle che riguardano il consenso informato. L'art. 6 del GDPR pone le basi giuridiche per il trattamento dei dati personali, evidenziando la necessità del "consenso" per avere la "liceità del trattamento". Sebbene tale articolo disponga che, il trattamento è "lecito" se vi è il "consenso dell'interessato" e il trattamento è "necessario all'esecuzione di un contratto di cui l'interessato è parte o all'esecuzione di misure precontrattuali adottate su richiesta dello stesso", nel caso dei data broker tale articolo non è rilevante poiché l'interessato non fa parte del contratto di scambio dei dati (i data broker scambiano i dati con terze parti e l'interessato non viene informato dello scambio) e inoltre il consenso dell'interessato non è trasferibile per altri scambi (se il soggetto fornisce il consenso ad esempio su una piattaforma online, tale consenso poi non può essere trasferito automaticamente se i dati vengono ceduti o riutilizzati da un altro soggetto, come i data broker). Inoltre, l'ottenimento di un "consenso informato" da parte dell'interessato (art. 6 e art.7 GDPR) si scontra con le pratiche dei data broker in quanto richiederebbe che questi ultimi informino con anticipo l'interessato sulla loro identità, sulle finalità del trattamento, sulle categorie di dati trattati, sulla profilazione, se i dati saranno poi rivenduti, sui destinatari del trattamento ecc. Quindi sebbene sia teoricamente possibile perseguire il modello di business dei data broker, in linea con le disposizioni del GDPR, ciò "richiederebbe condizioni preliminari per il consenso informato libero da parte dell'interessato a tutte le finalità per le quali i suoi dati saranno trattati". Tuttavia, i data broker trattano e vendono grandi volumi di dati, che riguardano una grande quantità di persone, per cui sarebbe logisticamente

difficile informare ogni persona sulle finalità e sui destinatari del trattamento ai fini di ottenere un consenso pienamente informato.

L'art. 21 del GDPR da all'interessato il diritto di opporsi al trattamento, quando le finalità del trattamento sono legate al "marketing diretto". Tuttavia, tale norma non è applicabile ai data broker, poiché essi non si occupano direttamente del marketing ma si concentrano sulla vendita di profili sui consumatori a terzi. I data broker non sono i titolari del trattamento, per cui per l'interessato diventa molto difficile far valere il suo diritto di opposizione nei loro confronti.

L'art. 6 stabilisce anche che il trattamento è lecito se persegue un "legittimo interesse" del titolare del trattamento o di terzi. Tuttavia, il concetto di "interesse legittimo" è definito in modo generico, includendo vari generi di interesse (legale, economico, idealista ecc.). Il solo "legittimo interesse" non limita tanto la base legale per il trattamento, per cui si sostiene nell'articolo che, l'interesse per essere considerato "legittimo" deve anche essere "necessario". Rimane argomento di dibattito come il GDPR valuti gli interessi meramente commerciali, in quanto né il GDPR né il diritto alla privacy, conferiscono all'individuo un diritto esclusivo alla commercializzazione dei propri dati. Nonostante dei modelli di business, orientati al profitto, siano legittimi e auspicabili da un punto di vista legale, il modello di business dei data broker sembra non possedere il requisito di "necessità", che renderebbe lecito il trattamento dei dati. Quindi basarsi sul "legittimo interesse" per giustificare lo scambio dei dati, a scopo di profitto, può risultare problematico. Infatti, è relativamente semplice e pratico dimostrare il "legittimo interesse", ma dimostrare la "necessità" del trattamento (che il trattamento sia essenziale per la finalità dichiarata) diventa più complicato. In generale si nota che gli interessi del titolare del trattamento o del responsabile, come in questo caso il data broker, sono opposti rispetto agli interessi dell'interessato, che guarda con molta attenzione il suo diritto alla protezione dei dati.

Come detto in precedenza, il requisito di "necessità" del trattamento è più complicato da dimostrare, anche perché è interpretato in modo molto restrittivo. Infatti, secondo la Corte di Giustizia dell'Unione europea, il trattamento è "necessario" solo quando il

“legittimo interesse” non può essere raggiunto in altri modi. Inoltre, secondo l’art. 5 i dati devono essere “minimizzati”, cioè il trattamento deve essere “limitato” a quanto necessario per raggiungere la finalità. Tuttavia, l’uso di algoritmi predittivi per fare inferenza sui consumatori e ottenere previsioni sempre più accurate è “intrinseco” nel modello di business dei data broker e va in contrasto col principio di “minimizzazione” dei dati. L’art. 5 prevede anche che i dati siano precisi e accurati e ciò non è tra gli interessi primari dei data broker. Inoltre, i dati devono essere trattati in un modo che risulti chiaro e trasparente per l’interessato; tuttavia, nel contesto dei data broker, è impossibile per l’interessato capire come tali dati vengono elaborati, in quanto si tratta di grandi volumi di dati. La mancanza di trasparenza nella raccolta e aggregazione dei dati va a intaccare ulteriormente i “diritti dell’interessato”, che così vede indebolirsi il controllo sui propri dati. Lo scambio di dati, col solo scopo di profitto, in generale non è conforme al GDPR. Tuttavia, il problema risiede nell’ “enforcement” del GDPR, in quanto gli interessati non possono far valere i propri diritti perché non sono a conoscenza delle pratiche dei data broker e di come i loro dati siano da loro raccolti, trattati, venduti, spesso in modo improprio.

L’anonimizzazione dei dati (obbligatoria per i dati personali, ai sensi del GDPR), spesso tirata in ballo dagli organismi regolatori, non sembra limitare il problema dell’asimmetria informativa tra interessati e data broker. Al contrario l’anonimizzazione porta gli interessati, che non sono consapevoli, a percepire un “falso senso di sicurezza”. Infatti, come accennato nei capitoli precedenti, i dati anche se anonimizzati possono essere “identificabili”, in quanto non sempre l’anonimizzazione del dato è totale.

Anche la trasparenza non basta a limitare gli eccessi negli scambi dei dati, dal momento che i data broker “non raggiungeranno mai una trasparenza significativa”. Anche una trasparenza totale non garantisce necessariamente una scelta consapevole da parte dell’utente. Al contrario la grande quantità di informazioni (informative sulle privacy e condizioni di utilizzo ad esempio) portano al “paradosso della privacy”, per cui la quantità di informazioni richiesta dagli utenti rende difficile e pesante la comprensione

di tali informazioni. Spesso gli utenti a causa della lunghezza delle informative sulla privacy accettano i termini senza aver compreso e letto il tutto.

Per quanto concerne le nuove direttive europee, il Data Act mira a garantire l'interoperabilità tra le piattaforme e stabilisce regole per un accesso equo e armonizzato ai dati generati dai dispositivi IoT per cittadini, imprese ed enti pubblici. Tuttavia, il suo campo di applicazione è limitato ed è difficile che possa avere degli effetti tangibili sul business dei data broker. Il Data Act ha obiettivi diversi rispetto alla protezione dei diritti dei cittadini, intaccati dalle pratiche dei data broker, e si concentra maggiormente sulla rimozione delle barriere all'ingresso per aumentare la competitività del mercato europeo, sulla promozione dell'innovazione e su strategie legate ad una sovranità digitale europea.

Il Data Governance Act stabilisce un quadro normativo chiaro per l'utilizzo dei dati nel settore pubblico. Ad ogni modo il DGA, non è di notevole importanza per il modello di business dei data broker, in quanto regola principalmente le condizioni per l'accesso ai dati da parte delle autorità pubbliche e per la condivisione volontaria dei dati ("altruismo dei dati"). Il DGA regolamenta i servizi di intermediazione di dati, quindi potenzialmente i data broker, ma non regolamenta la raccolta e la vendita dei dati a scopo di profitto, che è invece il modello tipico di business dei data broker. L'obiettivo del DGA è piuttosto di agevolare l'accesso e la condivisione dei dati tra pubblico e privato; pertanto, i data broker non hanno alcun incentivo ad operare sotto il quadro normativo del DGA, in quanto loro si occupano di raccolta su pervasiva e su larga scala di dati, nell'ottica del capitalismo di sorveglianza.

Invece i DSA e DMA non mirano esplicitamente a proteggere gli individui dalle pratiche dei data broker.

Il GDPR, sebbene tanto criticato e discusso, offre dei principi che potenzialmente possono contrastare il commercio improprio di dati, tuttavia è frenata da un deficit di "enforcement" che ne limita le potenzialità e rende difficile l'applicazione del regolamento, da parte dell'autorità pubblica, ai data broker. Inoltre, il GDPR tende a concentrarsi molto sulla raccolta dati sul browser (tramite cookie ad esempio), e meno

sulle tecnologie basate sulle applicazioni come i Software Development Kits (SDK)[24] (Reviglio, U., 2022), che permettono in ogni caso la raccolta dei dati da parte dei data broker (Morrison, 2020).

La nuova regolamentazione europea rappresenta un passo nella direzione giusta, ma manca di un quadro che regoli e limiti le asimmetrie informative e tecnologiche tra grandi raccoglitori ed elaboratori di dati, come i data broker e gli interessati. Una definizione chiara dei data broker e delle regole di trasparenza, come quelle richieste dal Digital Services Act per le "very large online platform" potrebbero essere dei buoni passi verso una maggiore regolazione dei data broker.

[24] Vedi Cap. 2 par. 2.2.1

CONCLUSIONE

La repentina crescita ed espansione del mercato dei dati e il ruolo sempre più centrale dei data broker sollevano questioni complesse e multidimensionali. I data broker possono potenzialmente apportare dei benefici ai consumatori, come ad esempio: i "prodotti per la prevenzione delle frodi" possono aiutarli a "prevenire che dei truffatori si spaccino per dei consumatori ignari", i "prodotti per il marketing" potrebbero consentire ai consumatori a trovare più facilmente i prodotti e i servizi di cui hanno bisogno o che preferiscono, i "people search product" potrebbero aiutare le persone a contattare "vecchi compagni di classe, vicini e amici" (FTC, 2014). Tuttavia, se questi prodotti contenessero errori o imprecisioni nei dati riportati potrebbero causare un'ingiusta esclusione dei consumatori da transazioni o servizi. Inoltre, la memorizzazione dei dati sensibili per periodi indefiniti aumenta il rischio di furto d'identità.

Va evidenziato che le opzioni di "opt-out" (ovvero la possibilità, fornita agli utenti, di richiedere ai data broker che i loro dati non siano trattati per determinati scopi ed eventualmente richiedere la cancellazione dei dati dai loro database) offerte dai data broker sono spesso poco visibili e limitate, rendendo difficile per i consumatori esercitare un pieno controllo sui propri dati (FTC, 2014).

La raccolta massiva e la vendita di informazioni personali, spesso sensibili, pongono significative riguardo la privacy e regolazione dei data broker. L'attuale quadro normativo europeo (GDPR e Big Data Acts) pur cercando di affrontare alcune problematiche, risulta spesso inadeguato a adattarsi alla complessità e specificità delle pratiche dei data broker, oltre ad avere difficoltà di enforcement. In questo contesto è fondamentale la "governance" dei dati, i quali, data la loro natura fluida, dinamica e multidimensionale, transitano tra diversi confini causando problemi di controllo e di giurisdizione. Pertanto, diventa necessario standardizzare le norme sulla governance a livello globale e non soltanto a livello europeo, come è stato fatto col Data Governance Act.

In futuro, una proposta di regolamentazione potrebbe consistere nella nomina di un'autorità sovranazionale e indipendente che si occupi della vigilanza sui data broker e sulle piattaforme digitali (Stiegler Center, 2019), oppure la creazione di principi internazionali condivisi assimilabili ad un "contratto sociale per i dati" (Reviglio, U., 2022). Secondo Cofone (2021) un passo in avanti potrebbe essere quello di rafforzare il principio della "limitazione della finalità" presente nel GDPR. Inoltre, data la complessità delle informazioni inferenziali prodotte da algoritmi e i potenziali rischi per i consumatori derivanti dalle inferenze, una proposta potrebbe essere quella di introdurre un "diritto alle inferenze ragionevoli" che obblighi i data broker a giustificare, ex ante, la ragionevolezza delle inferenze (Wachter & Mittelstadt, 2019).

In prospettiva di una futura regolazione dei data broker, sembra necessaria in primis, un'armonizzazione globale delle normative sulla protezione dei dati, unita ad un'agenda politica globale, ambiziosa e innovativa, capace di mitigare i rischi per la privacy e per i diritti dei cittadini e, al contempo, di favorire l'innovazione e il progresso tecnologico. Ciò sarà cruciale per costruire un futuro digitale più responsabile e sostenibile in cui i benefici economici della data economy siano suddivisi in modo equo tra tutti i cittadini.

BIBLIOGRAFIA

Acquisti Alessandro, Taylor Curtis, Wagman Liad. (2014). *The Economics of Privacy* in «Journal of Economic Literature»

Acquisti, A., Brandimarte, L., Loewenstein, G. (2015). *Privacy and human behavior in the age of information* in «Science»

AGCOM. (2018). *BIG DATA, Interim report nell'ambito dell'indagine conoscitiva di cui alla delibera n. 217/17/CONS*
https://www.agcom.it/sites/default/files/migration/rapporto/Studio-Ricerca%2008-06-2018.pdf

AGCM, AGCOM, Garante per la Protezione dei Dati Personali. (2020). *INDAGINE CONOSCITIVA SUI BIG DATA*
https://www.agcm.it/dotcmsdoc/allegati-news/IC_Big%20data_imp.pdf

Allen Marshall. (2018). *Health Insurers Are Vacuuming Up Details About You — And It Could Raise Your Rates* in «ProPublica»

Alovisio Mauro. (2016). *Regolamento ue 2016/679, ecco tutto ciò che cittadini e PA devono sapere* in «Agenda Digitale»
https://www.agendadigitale.eu/infrastrutture/nuovo-regolamento-privacy-ue-ecco-tutto-cio-che-cittadini-e-pa-devono-sapere/

American Civil Liberties Union. (2005). *Racial Profiling: Definition*
https://www.aclu.org/documents/racial-profiling-definition

Angwin Julia, Parris Terry Jr. (2016). *Facebook Lets Advertisers Exclude Users by Race* in «ProPublica»
https://www.propublica.org/article/facebook-lets-advertisers-exclude-users-by-race

Australian Competition and Consumer Commission's. (2023). *Digital Platform Services Inquiry – March 2024 report on data brokers*

Bagnoli Vicente. (2017). *The Definition of the Relevant Market, Verticalization and Abuse of Dominant Position in the Era of Big Data.*

Belfi, M. (2020). *California Consumer Privacy Act (CCPA): ambito di applicazione e regole di conformità per le aziende hi-tech*
https://www.cybersecurity360.it/legal/privacy-dati-personali/california-consumer-privacy-act-ccpa-ambito-di-applicazione-e-regole-di-conformita-per-le-aziende-hi-tech/

Belleflamme, P., Lam, W.M.W., and Vergote, W. (2020). *Competitive imperfect price discrimination and market power* in «Marketing Science»

Birckan Guilherme, Lima Dutra Moisés, De Macedo Douglas D. J., Godoy Viera Angel Freddy. (2020). *Personal data protection and its reflexes on the data broker industry*

Bounie David, Dubus Antoine, Waelbroeck Patrick. (2018). *Selling Strategic Information in Digital Competitive Markets.*

Bourreau Marc, De Streel Alexandre, Graef Inge. (2017). *Big Data and Competition Policy: Market Power, Personalised Pricing and Advertising*

Boyd, D., Crawford, K. (2012). *Critical questions for big data: Provocations for a cultural, technological, and scholarly phenomenon*

Cadwalladr Carole. (2017). *The great British Brexit robbery: how our democracy was hijacked* in «The Guardian»
https://www.theguardian.com/technology/2017/may/07/the-great-british-brexit-robbery-hijacked-democracy

Cadwalladr Carole , Graham-Harrison Emma. (2018). *Revealed: 50 million Facebook profiles harvested for Cambridge Analytica in major data breach* in «The Guardian»
https://www.theguardian.com/news/2018/mar/17/cambridge-analytica-facebook-influence-us-election

Calderini Barbara. (2021). *Programmatic advertising, così Google (e Facebook) manipolano il mercato degli annunci elettronici* in «Agenda Digitale»
https://www.agendadigitale.eu/mercati-digitali/programmatic-advertising-cosi-google-e-facebook-manipolano-il-mercato-degli-annunci-elettronici/

Carrascosa J.M., Mikians J., Cuevas R., Erramilli V., Laoutaris N. (2015). *I Always Feel Like Somebody's Watching Me Measuring Online Behavioural Advertising*

Casadesus-Masanell, R. and Hervas-Drane, A. (2015). *Competing with privacy* in «Management Science»

Cataleta A., Losavio A. (2022). *Digital Markets Act: cos'è e cosa prevede* in «Agenda Digitale»
https://www.agendadigitale.eu/mercati-digitali/digital-markets-act-cose-e-cosa-prevede/

Cataleta A., Longo A., Natale R. (2024). *GDPR, tutto ciò che c'è da sapere per essere in regola* in «Agenda Digitale» https://www.agendadigitale.eu/cittadinanza-digitale/gdpr-tutto-cio-che-ce-da-sapere-per-essere-preparati/

Cataleta, A., Losavio, A. (2024). *Data Act, il regolamento sull'accesso equo ai dati e sul loro uso: vantaggi e obblighi* in «Cyber Security 360» https://www.cybersecurity360.it/legal/data-act-ultimo-tassello-della-strategia-digitale-europea-vantaggi-e-obblighi/#_ftn1

Cataleta, A. (2023). *Data Governance Act ora applicativo: così cambia l'economia digitale*
https://www.agendadigitale.eu/sicurezza/privacy/la-data-economy-alla-prova-del-data-governance-act-lo-scenario/#_ftnref3

Cavanillas José María, Curry Edward, Wahlster Wolfgang Editors. (2016). *New Horizons for a Data-Driven Economy, A Roadmap for Usage and Exploitation of Big Data in Europe*

Christl Wolfie. (2017). *Corporate Surveillance In Everyday Life, How Companies Collect, Combine, Analyze, Trade, and Use Personal Data on Billions*

Citron Danielle Keats, Pasquale Frank. (2014). *The Scored Society: Due Process for Automated Predictions*

Clavorà Braulin Francesco ,Valletti Tommaso. (2016). *Selling customer information to competing firms* in «Economics Letters»
Cofone, I. (2021). *Beyond data ownership* in « Cardozo Law Review»

Condemi, J. (2022). *Digital Markets Act: cos'è e cosa prevede*
https://www.agendadigitale.eu/mercati-digitali/digital-markets-act-cose-e-cosa-prevede/

Condemi, J. (2022). *Digital Services Act: cos'è e cosa prevede la legge europea sui servizi digitali*
https://www.agendadigitale.eu/mercati-digitali/digital-services-act-cose-e-cosa-prevede-la-legge-europea-sui-servizi-digitali/

Conitzer, V., Taylor, C., and Wagman, L. (2012). *Hide and seek costly consumer privacy in a market with repeat purchases* in «Marketing Science»

Council of Economic Advisers. (2015). *Big Data and Differential Pricing*
https://obamawhitehouse.archives.gov/sites/default/files/whitehouse_files/docs/Big_Data_Report_Nonembargo_v2.pdf

Crain Matthew. (2018). *The limits of transparency: Data brokers and commodification*

De Mauro, Greco, Grimaldi. (2016). *A Formal Definition of Big Data Based on its Essential Features*

D'alessandro Jaime. (2018). *Scandalo Cambridge Analytica, parla Facebook: "I profili social coinvolti sono 87 milioni"* in «la Repubblica» https://www.repubblica.it/tecnologia/social-network/2018/04/04/news/scandalo_facebook-cambridge_analytica_i_profili_social_coinvolti_sono_87_milioni-192991515/

Della Piazza Sara. (2021). *Il caso Cambridge Analytica* in «DirittoConsenso» https://www.dirittoconsenso.it/2021/12/21/il-caso-cambridge-analytica/

Dijcks, J. (2013). *Oracle: Big data for the enterprise. Oracle White Paper*

DIRECTIVE 2006/123/EC OF THE EUROPEAN PARLIAMENT AND OF THE COUNCIL on services in the internal market. (2006). https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=CELEX:32006L0123

Di Resta Fabio. (2018). *Gdpr, la scelta del DPO: compiti e requisiti* in «Agenda Digitale» https://www.agendadigitale.eu/compiti-dpo

Duhigg Charles. (2007). *Bilking the Elderly, With a Corporate Assist* in «The New York Times» https://www.researchgate.net/publication/265225293_Bilking_the_Elderly_With_a_Corporate_Assist

Dumbill, E. (2013). *Making Sense of Big Data*

Equifax Annual Report. (2023). https://d1io3yog0oux5.cloudfront.net/_18cf1b1f549ee4879a576a94ae5052e1/equifax/db/2054/19426/annual_report/2023+Annual+Report.pdf

European Commission. (2018). *Consumer market study on online market segmentation through personalised pricing/offers in the European Union* https://commission.europa.eu/document/download/f6bae041-4ee7-443c-ad7b-f51f4ac14e10_en?filename=synthesis_report_online_personalisation_study_final.pdf

Experian Annual Report. (2016). https://www.experianplc.com/media/2744/discover-experian-fy17.pdf

Experian official website. (2024). https://www.experian.com.au/about-experian

Ezrachi, Ariel., Stucke Maurice E. (2016). *The rise of behavioural discrimination* https://wilte.wordpress.com/wp-content/uploads/2017/08/the-rise-of-behavioural-discrimination.pdf

Faroukhi Abou Zakaria, El Alaoui Imane, Gahi Youssef, Amine Aouatif. (2020). *Journal of Big Data, Big data monetization throughout Big Data Value Chain: a comprehensive review*

Federal Trade Commission. (2012). *Spokeo to Pay $800,000 to Settle FTC Charges Company Allegedly Marketed Information to Employers and Recruiters in Violation of FCRA* https://www.ftc.gov/news-events/news/press-releases/2012/06/spokeo-pay-800000-settle-ftc-charges-company-allegedly-marketed-information-employers-recruiters

Federal Trade Commission. *(2012). Protecting consumer privacy in an era of rapid change.* https://www.ftc.gov/sites/default/files/documents/reports/federal-trade-commission-report-protecting-consumer-privacy-era-rapid-change-recommendations/120326privacyreport.pdf

Federal Trade Commission. (2014). *Data Brokers, A Call for Transparency and Accountability.* https://www.ftc.gov/system/files/documents/reports/data-brokers-call-transparency-accountability-report-federal-trade-commission-may-2014/140527databrokerreport.pdf

Federal Trade Commission. (2016). *Big data: A tool for inclusion or esclusion? Understanding the issues (FTC Report)* https://www.ftc.gov/system/files/documents/reports/big-data-tool-inclusion-or-exclusion-understanding-issues/160106big-data-rpt.pdf

Google Flu Trends Estimates. (2008). https://www.google.com/publicdata/explore?ds=z3bsqef7ki44ac_#!ctype=l&strail=false&bcs=d&nselm=h&met_y=flu_index&scale_y=lin&ind_y=false&rdim=country&idim=country:ZA&ifdim=country&hl=en_US&dl=en_US&ind=false

Gu Yiquan, Madio Leonardo, Reggiani Carlo. (2021). *Data brokers co-opetition*

Hannak, A., Soeller, G., Lazer D., Mislove A., Wilson C. (2014*). Measuring Price Discrimination and Steering on E-commerce Web Sites* https://mislove.org/publications/Ecommerce-IMC.pdf

Hoofnagle Chris Jay. (2004). *Big Brother's Little Helpers: How ChoicePoint and Other Commercial Data Brokers Collect and Package Your Data for Law Enforcement* in «North Carolina Journal Of International Law»

Ichihashi, S. (2020). *Online privacy and information disclosure by consumers, American Economic Review.*

IDC. (2011). *IDC's worldwide big data taxonomy.*

Ikeda Scott. (2020). *Major Data Broker Exposes 235 Million Social Media Profiles in Data Leak* in «CPO Magazine»
https://www.cpomagazine.com/cyber-security/major-data-broker-exposes-235-million-social-media-profiles-in-data-leak/

Kaplan Levi, Mislove Alan, Sapieżyński Piotr. (2022). *Measuring Biases in a Data Broker's Coverage*

Kim Joanne. (2023). *Data Brokers and the Sale of Americans' Mental Health Data, The Exchange of Our Most Sensitive Data and What It Means for Personal Privacy* in «Duke Sanford Cyber Policy Program»
https://techpolicy.sanford.duke.edu/wp-content/uploads/sites/4/2023/02/Kim-2023-Data-Brokers-and-the-Sale-of-Americans-Mental-Health-Data.pdf

Kirchner Lauren. (2020). *When Zombie Data Costs You a Home* in « The Markup»

Knowledge Sourcing Intelligence. (2024). *Global Data Broker Market Size, Share, Opportunities, And Trends By Data Type (Consumer Data, Business Data), By End-User (BFSI, Retail, Automotive, Construction, Others), And By Geography – Forecasts From 2024 To 2029*
https://www.knowledge-sourcing.com/report/global-data-broker-market

Kosinski M., Stillwell D., Graepel T. (2013). *Private traits and attributes are predictable from digital records of human behavior* in «PNAS»
https://www.pnas.org/doi/epdf/10.1073/pnas.1218772110

Kuempel Ashley. (2016). The Invisible Middleman*: A Critique and Call for Reform of the Data Broker Industry* in «Northwestern Journal of International Law & Business»
Laney Doug. (2001). *3D Data Management: Controlling Data Volume, Velocity and Variety*
LexisNexis Risk Solutions website. (2024). https://risk.lexisnexis.com/about-us

Lonardo Francesca. (2017*). Data broker, chi sono e come evitano di violare la privacy* in «Agenda Digitale» https://www.agendadigitale.eu/cittadinanza-digitale/chi-sono-i-data-broker-e-come-evitare-di-violare-la-privacy/

Loukides Mike. (2010). *What is data science? The future belongs to the companies and people that turn data into products, O'Reilly Radar Report.*

Maietta Catia. (2018). *GDPR, i dodici nuovi diritti che i cittadini devono conoscere* in «Agenda Digitale» https://www.agendadigitale.eu/sicurezza/privacy/gdpr-guida-ai-diritti-del-cittadino/#_ftn1

Market Research Future. (2024). *Data Broker Market Research Report Information By Data Category (Consumer, Credit, Government, Technical, Real Estate, Education & Training, Product & Services, Risk Management, Data Types), Data Type (Unstructured Data, Structured Data, and Custom Structure Data), Pricing Model (Subscription Paid, Pay Per Use Paid and Hybrid Paid Models), End Use Sector (BFSI, Retail And FMCG, Manufacturing, Media, Government Sector, and Others Sector), And By Region (North America, Europe, Asia-Pacific, And Rest of The World) – Forecast Till 2032*. https://www.marketresearchfuture.com/reports/data-broker-market/toc

Maximize Market Research. (2024). Data Broker Market is expected to grow CAGR of 7.25% Over 202 4-30, this Report Covers Global Analysis by Data Category, Data Type, End User, Growth and Forecast (2024-2030) | MMR https://www.maximizemarketresearch.com/market-report/global-data-broker-market/55670/

Menietti Emanuele. (2018). *Il caso Cambridge Analytica, spiegato bene* in «Il Post» https://www.ilpost.it/2018/03/19/facebook-cambridge-analytica

Montes, R., Sand-Zantman, W., Valletti, T. (2019). *The value of personal information in online markets with endogenous privacy* in «Management Science»

Morrison, S. (2020). *The hidden trackers in your phone, explained: How covert code enables your phone's apps to spy on you.* https://www.vox.com/recode/2020/7/8/21311533/sdks-tracking-data-location

Muralidhar Krishnamurty, Palk Laura. (2018). *A Free Ride: Data Brokers'Rent-Seeking Behavior and the Future of Data Inequality* in «Vanderbilt Journal of Entertainment & Technology Law»

National Institute of Standards and Technology. (2014*). Big Data Public Working Group. Draft of Big Data Definition.*

Newcomer, E. (2017). *Uber Starts Charging What It Thinks You're Willing to Pay* in «Bloomberg» https://www.bloomberg.com/news/articles/2017-05-19/uber-s-future-may-rely-on-predicting-how-much-you-re-willing-to-pay

Niola, F. (2024). *Il Data Act ridefinisce la sovranità informativa UE: i chiarimenti* in «Agenda Digitale» https://www.agendadigitale.eu/sicurezza/privacy/data-act-la-condivisione-forzata-dei-dati-minaccia-privacy-e-sovranita-digitale/

OECD (Organisation for Economic Co-operation and Development). (2018). *Personalised Pricing in the Digital Era*
https://one.oecd.org/document/DAF/COMP(2018)13/en/pdf

O'Neil Cathy. (2016). *Weapons of Math Destruction*

Otto Paul N., Antón Annie I., Baumer David L. (2007). *The ChoicePoint Dilemma: How Data Brokers Should Handle the Privacy of Personal Information* in « IEEE Security & Privacy»

Parham Jason. (2018). *Targeting Black Americans, Russia's IRA Exploited Racial Wounds* in «Wired»
https://www.wired.com/story/russia-ira-target-black-americans/

Pigou, A.C. (1920). *The economics of welfare.*

Polimeni Antonino. (2022). *Cookie: cosa sono, come funzionano e come proteggerti* in «Agenda Digitale»
https://www.agendadigitale.eu/infrastrutture/tutto-quello-che-dobbiamo-sapere-sui-cookie-per-la-privacy-da-utenti-o-gestori/

Politini, S. (2020). *Diventare una data-driven company: l'esperienza di Cattolica Assicurazioni*
https://www.digital4.biz/executive/digital-transformation/cattolica-assicurazioni-diventa-una-data-driven-company/

Popper Ben. (2014). *How the NYPD is using social media to put Harlem teens behind bars, The untold story of Jelani Henry, who says Facebook likes landed him in Rikers* in «The Verge»
https://www.theverge.com/2014/12/10/7341077/nypd-harlem-crews-social-media-rikers-prison

Porter Michael E. (1985). *Competitive Advantage: Creating and Sustaining Superior Performance*

Proposta di REGOLAMENTO DEL PARLAMENTO EUROPEO E DEL CONSIGLIO relativo al rispetto della vita privata e alla tutela dei dati personali nelle comunicazioni elettroniche e che abroga la direttiva 2002/58/CE (regolamento sulla vita privata e le comunicazioni elettroniche). (2017).
https://eur-lex.europa.eu/legal-content/IT/TXT/PDF/?uri=CELEX:52017PC0010

REGOLAMENTO (UE) 2016/679 DEL PARLAMENTO EUROPEO E DEL CONSIGLIO del 27 aprile 2016 relativo alla protezione delle persone fisiche con riguardo al trattamento dei dati personali, nonché alla libera circolazione di tali dati e

che abroga la direttiva 95/46/CE (regolamento generale sulla protezione dei dati) (2016). https://eur-lex.europa.eu/legal-content/IT/TXT/PDF/?uri=CELEX:32016R0679

REGOLAMENTO (UE) 2018/302 DEL PARLAMENTO EUROPEO E DEL CONSIGLIO del 28 febbraio 2018 recante misure volte a impedire i blocchi geografici ingiustificati e altre forme di discriminazione basate sulla nazionalità, sul luogo di residenza o sul luogo di stabilimento dei clienti nell'ambito del mercato interno e che modifica i regolamenti (CE) n. 2006/2004 e (UE) 2017/2394 e la direttiva 2009/22/CE (2018). https://eur-lex.europa.eu/legal-content/IT/TXT/PDF/?uri=CELEX:32018R0302

REGOLAMENTO (UE) 2022/2065 DEL PARLAMENTO EUROPEO E DEL CONSIGLIO del 19 ottobre 2022 relativo a un mercato unico dei servizi digitali e che modifica la direttiva 2000/31/CE (regolamento sui servizi digitali). (2022). https://eur-lex.europa.eu/legal-content/IT/TXT/PDF/?uri=CELEX:32022R2065

REGOLAMENTO (UE) 2022/1925 DEL PARLAMENTO EUROPEO E DEL CONSIGLIO del 14 settembre 2022 relativo a mercati equi e contendibili nel settore digitale e che modifica le direttive (UE) 2019/1937 e (UE) 2020/1828 (regolamento sui mercati digitali). (2022). https://eur-lex.europa.eu/legal-content/IT/TXT/PDF/?uri=CELEX:32022R1925

REGOLAMENTO (UE) 2022/868 DEL PARLAMENTO EUROPEO E DEL CONSIGLIO del 30 maggio 2022 relativo alla governance europea dei dati e che modifica il regolamento (UE) 2018/1724 (Regolamento sulla governance dei dati). (2022). https://eur-lex.europa.eu/legal-content/IT/TXT/PDF/?uri=CELEX:32022R0868

Reinsel David, Gantz John, Rydning John, IDC. (2018). *Data Age 2025 The Digitization of the World From Edge to Core*

Rieke, A., Yu, H., Robinson, D., Van Hoboken, J. University of Amsterdam. (2016). *Data Brokers in an Open Society*

Ruschemeier Hannah. (2023). *Data Brokers and European Digital Legislation*

Schroeck, M., Shockley, R., Smart, J., Romero-Morales, D., & Tufano, P., *NY: IBM* Institute for Business Value, Said Business School. (2012). *Analytics: The real-world use of big data. New York*

Scott Mark. (2018). *Cambridge Analytica helped 'cheat' Brexit vote and US election, claims whistleblower* in «Politico»

https://www.politico.eu/article/cambridge-analytica-chris-wylie-brexit-trump-britain-data-protection-privacy-facebook/

Shafer, T. (2017). *The 42 V's of Big Data and Data Science*

Sherman Justin, Hoffman David, Reeves Spencer, Klein Aden, Allen Kruse Brady, Simmons Alistair, Barton Hayley. (2023). *Response from Duke University's Data Brokerage Research Project, Consumer Financial Protection Bureau (CFPB) Request for Information Regarding Data Brokers and Other Business Practices Involving the Collection and Sale of Consumer Information*

Sherman Justin. (2021). *Data Brokers and Sensitive Data on U.S. Individuals, Threats to American Civil Rights, National Security, and Democracy*
Shy Oz, Stenbacka Rune. (2016). *Customer Privacy and Competition* in «Journal of Economics & Management Strategy»

Sorte Simone. (2020). *Facebook, lo scandalo Cambridge Analytica spiegato in modo semplice* in «Digital Flow» https://digitalflow.it/scandalo-facebook-cambridge-analytica/#Cose_Cambridge_Analytica

STATISTA. (2021). *Volume of data/information created, captured, copied, and consumed worldwide from 2010 to 2020, with forecasts from 2021 to 2025* https://www.statista.com/statistics/871513/worldwide-data-created/
Stevens, Gina Marie. (2007). *Data brokers: Background and industry overview* in «CRS Report for Congress»

Stigler Center for the Study of the Economy and the State. (2019). *Stigler Committee on Digital Platforms Final Report*
https://www.chicagobooth.edu/-/media/research/stigler/pdfs/digital-platforms---committee-report---stigler-center.pdf

The Associated Press. (2021). *Priest outed via Grindr app highlights rampant data tracking* in «NBC News»
https://www.nbcnews.com/tech/security/priest-outed-grindr-app-highlights-rampant-data-tracking-rcna1493

Thompson Stuart A., Warze Charlie. (2019). *How to Track President Trump* in «New York Times»
https://www.nytimes.com/interactive/2019/12/20/opinion/location-data-national-security.html

Tobin Ariana, Kofman Ava. (2022). *Facebook Finally Agrees to Eliminate Tool That Enabled Discriminatory Advertising* in «ProPublica»
https://www.propublica.org/article/facebook-doj-advertising-discrimination-settlement

TransUnion Annual Report. (2023).
https://investors.transunion.com/~/media/Files/T/Transunion-IR-V2/annual-reports/2023/2023-annual-report.pdf

United States Senate, Committee On Commerce, Science, And Transportation, Office Of Oversight And Investigations Majority Staff. (2013). *A Review of the Data Broker Industry: Collection, Use, and Sale of Consumer Data for Marketing Purposes* in «Staff Report For Chairman Rockefeller»
https://www.commerce.senate.gov/services/files/0d2b3642-6221-4888-a631-08f2f255b577

U.S Department of Justice, Office of Public Affairs. (2016). *Justice Department and Consumer Financial Protection Bureau Reach Settlement to Resolve Allegations of Auto Lending Discrimination by Toyota*
https://www.justice.gov/opa/pr/justice-department-and-consumer-financial-protection-bureau-reach-settlement-resolve-0

Valsania Marco. (2019). *Facebook, multa record da 5 miliardi. Violata la privacy nel caso Cambridge Analytica* in «Il Sole 24 Ore»
https://www.ilsole24ore.com/art/facebook-multa-record-5-miliardi-violata-privacy-caso-cambridge-analytica-ACzEObY

Varian Hal R. (1987). *Price Discrimination*

Venkatadri Giridhari, Sapiezynski Piotr, Redmiles Elissa M., Mislove Alan, Goga Oana, Mazurek Michelle L., Gummadi Krishna P. (2019). *Auditing Offline Data Brokers via Facebook's Advertising Platform* in «HAL Open Science»
https://hal.science/hal-02069470/file/databrokers-measurement_finalCameraReady.pdf

Verisk Annual Report, Experience the Possible. (2023).
https://s29.q4cdn.com/767340216/files/doc_financials/2023/ar/verisk-2023-annual-report.pdf

Wachter, S., Mittelstadt, B. (2019). *A Right to Reasonable Inferences: Re-Thinking Data Protection Law in the Age of Big Data and AI* in « Columbia Business Law Review »

Warren, S., Brandeis, L. (1890). *The Right to Privacy* in «Harvard Law Review»

Westin, A. (1967). *Privacy And Freedom* in «Washington and Lee Law Review»

Zanotti, L. (2024). *Smart manufacturing: cos'è e come funziona una fabbrica connessa*
https://www.digital4.biz/supply-chain/smart-manufacturing-cose-fabbrica-connessa-funzionamento/

Zuboff Shoshana. (2019). *The Age of Surveillance Capitalism, The Fight for a Human Future at the New Frontier of Power*

SITOGRAFIA

https://www.illion.com.au/who-we-are/

https://www.corelogic.com/why-corelogic/

https://www.verisk.com/company/about/

https://www.nielsen.com/about-us/locations/australia/

https://www.digital4.biz/marketing/advertising/cookie-di-terza-parte-e-prima-parte-cosa-sono-come-cambia-la-pubblicita-online/

https://www.infodata.ilsole24ore.com/2022/01/02/quanti-dati-generati-minuto/#:~:text=Nel%202020%2C%20sono%20stati%20creati,di%20180%20Zb%20nel%202025.

https://www.ftc.gov/sites/default/files/documents/cases/2008/08/080801reeddo.pdf

https://www.ilsole24ore.com/art/il-crollo-facebook-67percento-scia-scandalo-cambridge-pesa-wall-street--AEiHkDJE

https://www.garanteprivacy.it/web/guest/home/docweb/-/docweb-display/docweb/9121352

https://tg24.sky.it/mondo/2018/04/04/facebook-cambridge-analytica-87-milioni-profili-coinvolti

https://www.iubenda.com/it/help/5424-guida-gdpr

https://www.unolegal.it/guida-gdpr/

https://www.garanteprivacy.it/home/principi-fondamentali-del-trattamento

https://www.garanteprivacy.it/valutazione-d-impatto-della-protezione-dei-dati-dpia-

https://www.sap.com/italy/products/artificial-intelligence/what-is-iot.html

https://www.agendadigitale.eu/sicurezza/gdpr-sanzioni-e-responsabilita-tutto-cio-che-ce-da-sapere/#Le_sanzioni_amministrative_GDPR

https://www.agendadigitale.eu/sanita/ehds-verso-lunione-sanitaria-europea-cose-le-cautele-i-vantaggi-per-i-cittadini/

https://digitexport.promositalia.camcom.it/informazione/normativa/il-digital-markets-act-cos-e-e-cosa-prevede.kl#:~:text=Il%20%22Digital%20Markets%20Act%22%20(,ovvero%20quelle%20che%20detengono%20una

https://www.altalex.com/documents/news/2024/07/11/regolamentazione-digitale-digital-services-act-piattaforme-online

https://www.cybersecurity360.it/legal/privacy-dati-personali/data-governance-act-nuove-opportunita-e-rischi-della-data-economy-europea/

https://iabeurope.eu/press-release-european-online-advertising-surpasses-tv-to-record-annual-spend-of-e36-2bn/

https://www.rockwellautomation.com/content/dam/rockwell-automation/documents/pdf/campaigns/state-of-smart-2024/9th-annual-state-of-smart-manufacturing-report-en.pdf

https://www.cybersecurity360.it/legal/digital-services-act-e-riforma-del-panorama-digitale-la-stretta-sulle-big-tech/

https://www.agendadigitale.eu/mercati-digitali/digital-service-act-perche-le-aziende-italiane-devono-prepararsi-al-17-febbraio-2024/

https://ec.europa.eu/commission/presscorner/detail/en/ip_23_2413

https://www.consilium.europa.eu/it/press/press-releases/2021/11/25/what-is-illegal-offline-should-be-illegal-online-council-agrees-on-position-on-the-digital-services-act/

https://digital-strategy.ec.europa.eu/it/policies/dsa-vlops#:~:text=Il%20DSA%20classifica%20le%20piattaforme,dimensioni%20molto%20grandi%20(VLOSE).

https://www.ftc.gov/legal-library/browse/statutes/fair-credit-reporting-act