



**Politecnico
di Torino**

Politecnico di Torino

Ingegneria Gestionale – Gestione dell’Innovazione e Imprenditorialità

A.a. 2024/2025

Sessione di Laurea Ottobre 2025

Il Mercato dei Big Data

Con riferimento al Ruolo dei Personal Data Brokers

Relatori:

Carlo Cambini

Candidati:

Mattia Fenoglio

Abstract

L'elaborato cerca di analizzare il mercato dei Big Data, con particolare attenzione al ruolo dei Personal Data Broker e alle principali soluzioni da essi proposte.

Nel capitolo 1 è approfondito il concetto di dato, andando ad analizzare il ciclo di vita, il valore e le diverse tipologie. Si pone attenzione all'evoluzione delle tecnologie digitali, le quali hanno favorito una crescita esponenziale nell'utilizzo e nella gestione del dato.

Il capitolo 2 presenta una panoramica sulle dimensioni del mercato dei Big Data, soffermandosi sugli attori coinvolti e sulle operazioni di fusione e acquisizione che ne caratterizzano lo sviluppo. In questa sezione trova spazio anche l'analisi del modello elaborato da Jan Krämer, Daniel Schnurr e Bastian Haberer, integrato da esempi attuali e da considerazioni personali che tendono a ridimensionare il ruolo dei Personal Data Broker all'interno del mercato.

Nel capitolo 3 si è proceduto concentrandosi sui principali Personal Data Broker a livello internazionale. Oltre a descrivere le soluzioni e la tipologia di dati raccolti si è cercato di trovare, con difficoltà, una valutazione economica dei servizi offerti.

Infine, il capitolo 4 presenta un caso antitrust in cui il dato assume un ruolo centrale.

1 – I BIG DATA – INTRODUZIONE	4
1.1 Dato – definizione	4
1.2 Ciclo di vita	4
1.3 Big Data – definizione	8
1.4 Storia	9
1.5 Caratteristiche	10
1.6 Fonti	12
1.7 Big Data Value Chain	13
1.8 Tecnologie abilitanti	16
1.9 Tipologie di dati.....	19
2 – IL MERCATO DEI BIG DATA	22
2.1 Introduzione al mercato.....	22
2.2 Le dimensioni del mercato.....	24
2.2.1 In Italia.....	25
2.3 Gli attori in generale	26
2.3.1 Consumatori.....	26
2.3.2 Content and Service Providers – CSP	27
2.3.3 Personal Information Management Systems – PIMS.....	27
2.3.4 Personal Data Brokers – PDB	27
2.3.5 Regolatori e policy makers	28
2.3.6 Esempi pratici	29
2.4 Approfondimento sui Personal Data Brokers	29
2.4.1 Modello base senza PDB.....	31
2.4.2 Modello base con PDB	31
2.4.3 Assunzioni del modello base.....	32
2.4.3 Considerazioni sui consumatori.....	32
2.4.4 Considerazioni sui Content and Service Provider (CSP).....	34
2.4.5 Considerazioni sui Personal Data Broker (PDB)	35
2.4.6 Equilibrio senza PDB	36
2.4.7 Commento al modello senza PDB	36
2.4.8 Equilibrio con PDB	37

2.4.9 Riassunto risultati matematici del modello con PDB.....	43
2.4.10 Commento del modello con PDB.....	44
2.5 I principali attori nel mercato	48
2.6 Una tendenza del mercato: le acquisizioni.....	50
2.7 I dati raccolti:	53
2.7.1 Marketing	56
2.7.2 Valutazioni creditizie	57
2.7.3 Informazioni mediche	59
3 – Personal Data Brokers	60
3.1 Acxiom.....	60
3.2 Circana.....	62
3.3 Merkle	65
3.4 Lenddo	68
3.5 DM Databases	73
4 – CASO ANTITRUST	80
4.1 Introduzione al caso	80
4.2 Il caso	80
4.3 L'accusa	82
4.4 Il ruolo del dato	83
4.5 Accenno di Teoria	84
4.6 Commento	84
Conclusione.....	86
Bibliografia e Sitografia:.....	88

1 – I BIG DATA – INTRODUZIONE

1.1 Dato – definizione

Il dato, che deriva dal latino datum, è una rappresentazione codificata (numerica o simbolica) di un'entità, di un fenomeno, di una transazione o di una caratteristica. È una sequenza di semplici fatti che rappresenta degli eventi accaduti all'interno di un'organizzazione (ad esempio una transazione commerciale) o semplicemente nel mondo fisico, prima che siano stati organizzati e disposti in una forma che possa essere compresa e utilizzata dalle persone. È importante che ai dati sia attribuita una forma strutturata e un chiaro significato in modo da poter aiutare gli esseri umani, ad esempio, nei processi decisionali all'interno di un'organizzazione.

Si vuole specificare che il dato si discosta dall'informazione in quanto essa possiede già un significato ed è associata ad un determinato contenuto.

Il dato per poter essere di qualità, in accordo con lo Standard ISO/IEC 25012:2008 deve presentare almeno le seguenti caratteristiche:

- Accuratezza: perfetta corrispondenza con il valore reale del concetto a cui si riferisce.
- Coerenza: non deve presentare contraddittorietà rispetto ad altri dati nel contesto d'uso.
- Completezza: deve risultare esaustivo, presentando tutti gli attributi necessari.
- Attualità: aggiornato al tempo rispetto al momento in cui è utilizzato.

1.2 Ciclo di vita

La gestione e la comprensione del ciclo di vita dei dati è essenziale e utile per qualsiasi organizzazione affinché essi siano utilizzati ed elaborati per produrre analisi accurate e ottimizzazione delle pratiche di gestione.

Il ciclo di vita del dato, come riportato nella figura sottostante, può essere suddiviso in 8 fasi che analizzeremo nel dettaglio:

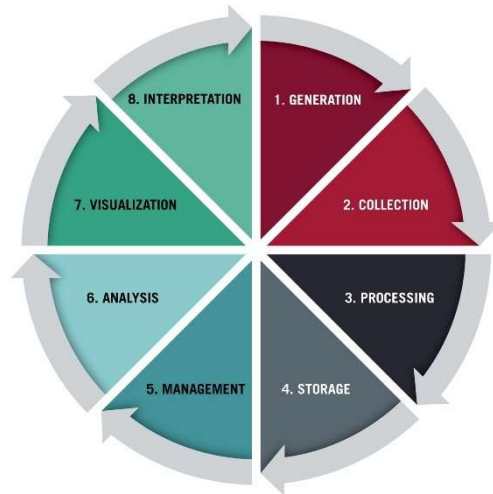


Figura 1: 8 steps in the data life cycle - Harvard Business School

- **Generazione:** perché il ciclo di vita dei dati possa iniziare essi devono essere generati. Essi sono creati da una varietà di fonti (es. interazioni con i clienti, transazioni aziendali, dispositivi dell'IoT, attività sui social media...). È importante sottolineare che questa generazione avviene indipendentemente dal fatto che l'utente o il consumatore ne sia consapevole o meno, soprattutto in un mondo come il nostro che tende sempre più all'online.
- **Raccolta:** è importante che i dati siano aggregati in maniera accurata. Essi devono essere raccolti in maniera strutturata evitando perdite eccessive ma allo stesso tempo eliminando quelli inutili o irrilevanti. I dati possono essere raccolti tramite sondaggi, interviste, osservazione diretta, registri delle transazioni, interfacce di programmazione delle applicazioni...
- **Elaborazione:** una volta che i dati sono raccolti è necessario che siano puliti (rimozione di duplicati, completamento di valori mancanti...), trasformati in un formato più strutturato che possa essere archiviato in maniera più efficiente e infine crittografati così da essere protetti da problemi di privacy; questi passaggi garantiscono una maggiore accuratezza e precisione dei dati.
- **Archiviazione:** questa fase è essenziale per garantire che i dati siano accessibili e sottoposti a procedure di backup per un utilizzo futuro.

Questo avviene per mezzo della creazione di database che possono essere successivamente archiviati in cloud, server o disco rigido. È importante scegliere la giusta soluzione di archiviazione per permettere una protezione idonea, una migliore organizzazione, un utilizzo efficiente ed una più facile scalabilità.

- **Gestione:** questa fase, è bene che non sia considerata come una fase statica. Si tratta infatti di un processo continuo che presiede l'intero progetto di analisi dati e comporta la loro continua organizzazione e manutenzione. Avviene in diverse maniere: tramite la definizione di standard e dei ruoli degli utenti (chi può accedere a cosa), attraverso il continuo monitoraggio e l'implementazione di crittografia, controlli di accesso ed audit (strumento per individuare vulnerabilità e anomalie). È importante anche che sia creata una visione unificata dei dati a garanzia di una coerenza generale. Infine, è di fondamentale importanza che siano implementate politiche standard che mantengano la privacy dei dati.
- **Analisi:** in questa fase si cerca di ricavare intuizioni significative dai dati grezzi. Si utilizzano strumenti e tecniche che servono a identificare modelli, tendenze e correlazioni. Tra i metodi più utilizzati si vuole citare: la modellazione statistica, gli algoritmi, l'intelligenza artificiale, il data mining e il machine learning. Questa fase è molto importante in quanto permette, ad esempio alle aziende, di prendere delle decisioni più informate.
- **Visualizzazione:** in questa fase i dati diventano più comprensibili attraverso, ad esempio, grafici e diagrammi, dashboard interattive, mappe, scatter plot, istogrammi e tree map. Questi strumenti permettono di comunicare più rapidamente l'analisi effettuata ad un pubblico più ampio sia all'interno che all'esterno dell'organizzazione.
- **Interpretazione:** qui i dati visualizzati ed analizzati sono utilizzati per prendere decisioni in maniera informata. È importante dare senso ai risultati analitici e trarre conclusioni per poter spiegare le performance

aziendali. È altrettanto di fondamentale rilevanza fornire indicazioni strategiche, ad esempio, sul marketing o sviluppo prodotto o coinvolgimento del cliente. Questa fase dovrebbe assicurare che tutte le intuizioni derivate dall'analisi e dalla visualizzazione dei dati siano efficacemente utilizzate per guidare decisioni strategiche e migliorare i risultati dell'organizzazione.

Si ribadisce che un'attenta comprensione del ciclo di vita dei dati permetta di comunicare in maniera più efficace gli obiettivi e i successivi piani per un'organizzazione. Questo consente, infatti, un miglioramento dei processi decisionali in quanto, processi con dati strutturati, portano a intuizioni più affidabili a cui susseguono decisioni strategiche informate. Inoltre, un punto da sottolineare è che un ciclo di vita dei dati bene definito permette alle organizzazioni di scalare in maniera efficiente le operazioni man mano che le esigenze aumentano garantendo loro flessibilità e possibilità di adattarsi a nuove fonti dati e a tecnologie innovative.

Infine, per una maggiore completezza si riporta il ciclo di vita dei dati con una breve spiegazione per ciascuna fase:

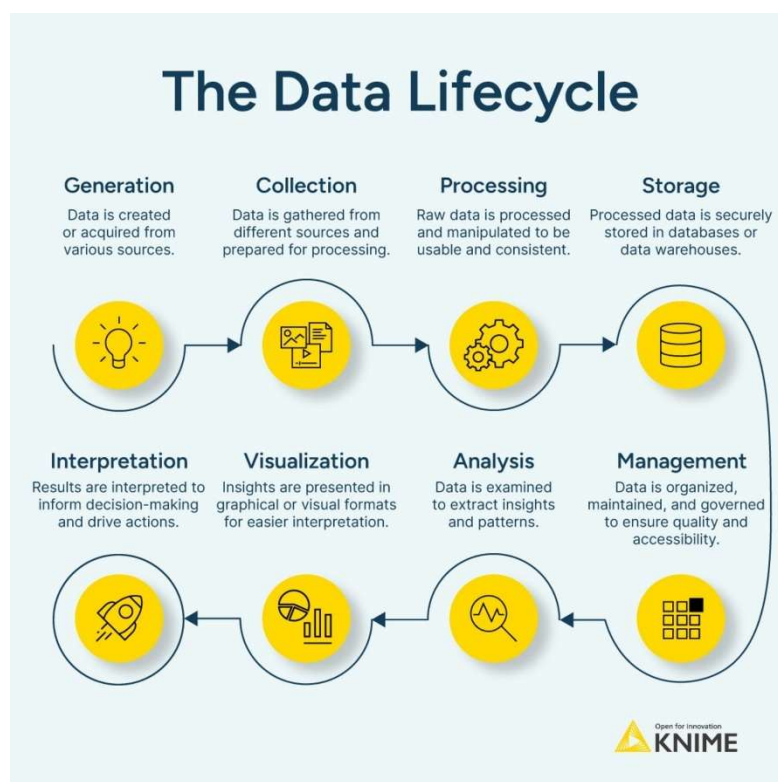


Figura 2: Data lifecycle: The 8 stages and who is involved - KNIME

1.3 Big Data – definizione

Oggi giorno, con l'avvento della rete web, dell'IoT, delle tecnologie NFC (Near Field Communication) e di moltissime altre fonti digitali la quantità di dati prodotta è cresciuta in modo esponenziale. Basta pensare che all'inizio del 2020 il World Economic Forum (WEF) ha stimato una quantità di dati prodotti pari a 44 zettabytes ($44 * 10^{21}$) per poi raggiungere, entro la fine del 2025 i 175 ZB¹: si parla infatti di Big Data.

I Big Data si riferiscono a set di dati estremamente vasti e complessi che non possono essere facilmente gestiti o analizzati con gli strumenti di elaborazione dei dati tradizionali (es. fogli Excel o database relazionali) indipendentemente dalla potenza di calcolo o dalla capacità di archiviazione fisica disponibile: necessitano dunque di nuove tecnologie.

I Big Data possono essere prodotti dalle persone (es. attraverso i social media), dai processi (es. transazioni commerciali) oppure possono essere generati dalle macchine le quali li raccolgono attraverso varie tipologie di sensori ad esempio dispositivi collegati all'IoT (GPS, frigoriferi, auto ecc. ecc.).

La Commissione Europea ha definito i Big Data con l'immagine riportata sotto: come descritto precedentemente essi provengono da diverse fonti e sono analizzati per ottenere informazioni utili in numerosi campi tra cui, ad esempio, medico, agricoltura, settore pubblico e trasporti e industriale.

Si specifica infatti che i Big Data, una volta raccolti ed analizzati, sono utilizzati dalle organizzazioni, sia aziendali che non, con intenti di mercato e di profilazione delle abitudini, dei comportamenti e delle caratteristiche degli utenti, cittadini e clienti.

Questo utilizzo, come emerge dalle analisi dei capitoli successivi, non sempre avviene in modo trasparente.

¹ Global DataSphere Forecast (IDC)



Figura 3: Big data: definizione, benefici e sfide - Parlamento europeo

1.4 Storia

Il termine Big Data è utilizzato sin dall'inizio degli anni '90. Sebbene non si sappia esattamente chi sia stato il primo a coniare il termine, molti attribuiscono a John R. Mashey, informatico e imprenditore statunitense che all'epoca lavorava alla Silicon Graphics (azienda informatica statunitense), il merito di averlo reso popolare.

In realtà, nella loro essenza, i Big Data non sono qualcosa di completamente nuovo o di recente introduzione. Infatti, fin dagli albori della civiltà, l'uomo ha cercato di utilizzare i dati per migliorare il processo decisionale o per ottenere un vantaggio competitivo (per lo più militare).

Tra i primi esempi possiamo citare la Biblioteca di Alessandria, fondata intorno al 300 a.C., che può essere considerata il primo tentativo degli antichi egizi di acquisire tutti i dati all'interno dell'impero: si stima che la biblioteca fosse composta da 40.000 a 400.000 papiri (l'equivalente di circa 100.000 libri)². Persino gli antichi leader del mondo si resero conto che combinare diverse fonti di dati avrebbe potuto portare a un vantaggio sugli altri imperi concorrenti.

² Wiegand, W.A and Donald Jr, G., 2015 Encyclopedia of library history (Vol. 503) Routledge

Altri casi d'uso ben documentati delle prime forme di raccolta e analisi dati provengono dall'Impero Romano: gli antichi militari romani utilizzavano analisi statistiche molto dettagliate per "prevedere" a quale confine sarebbe stata più probabile un'insurrezione nemica così da poter organizzare la distribuzione ottimale delle proprie risorse militari.

Curioso notare inoltre come un altro esempio di grande raccolta dati (che oggi potremmo chiamare Big Data) risalga anch'esso all'Antica Roma: il Censo Romano. Esso fu istituito da Servio Tullio (dal 578 a.C. al 534 a.C.) con lo scopo di creare un elenco dei cittadini e dei loro beni nell'antica Roma. Tutti i dati raccolti servivano a stabilire quali cittadini erano obbligati a prestare il servizio militare e a quali condizioni, e ovviamente a determinare gli obblighi fiscali. I romani, data la vastità dell'impero e quindi la quantità enorme di dati raccolti si resero conto immediatamente di come fosse impossibile procedere nell'elaborazione dei dati in modo centralizzato adottarono dunque sistemi distribuiti; in maniera concettualmente simile, oggi, l'enorme mole di dati digitali è spesso gestita con architetture distribuite che permettono di elaborare le informazioni vicino alla loro fonte accelerando così i tempi di risposta.

Attualmente gli aspetti chiave sono la disponibilità e l'accessibilità a enormi quantità di dati. Mentre fino agli anni '50 la maggior parte delle analisi dei dati era eseguita manualmente e su carta, ora disponiamo della tecnologia e della capacità di analizzare terabyte di dati in frazioni di secondo. Dall'inizio del XXI secolo, il volume e la velocità con cui i dati sono generati e analizzati sono cambiati in modo quasi incomprensibile alla specie umana.

1.5 Caratteristiche

Inizialmente essi erano caratterizzati da un modello a 3V:

- Volume: i Big Data hanno dimensioni enormi, che li porta a non poter essere gestiti con i tradizionali strumenti di database. È possibile stimare il volume dei dati generati da un'azienda all'ordine di grandezza del terabyte (10^{12}) o del petabyte (10^{15}). Si vuole far notare come il volume di dati a livello mondiale stia crescendo a un tasso annuo del 60% e per di più il 90% dei dati esistenti oggi nel mondo è stato creato negli ultimi 2 anni. ³

³ Data and Information Quality di Carlo Batini e Monica Scannapieco

- Velocità: i Big Data sono generati molto rapidamente e per di più la maggior parte delle volte sono elaborati in tempo reale: ogni minuto sono pubblicati 400.000 tweet su Twitter, inviati 200 milioni email e sottomesse 2 milioni di query di ricerca su Google.⁴
- Varietà: un set di Big Data in genere contiene dati strutturati, semi-strutturati e non strutturati. Spesso sono caratterizzati dall'assenza di una strutturazione definita e da documenti di vario genere (txt, csv, PDF, Word...), blog post, commenti social. Infine, anche la provenienza di questi Big Data è molto variegata: alcuni sono generati automaticamente da macchine, altri provenienti da sensori oppure altri ancora prodotti dagli utenti del web.

A questo modello iniziale delle 3V sono state aggiunte altre 2V:

- Veridicità: è molto importante che il valore informativo estraibile dai Big Data sia veritiero. La loro accuratezza/veridicità è estremamente importante in quanto lievi anomalie, distorsioni o rumori possono causare ricadute ad esempio nelle scienze algoritmiche.
- Valore: è importante riuscire ad estrarre informazioni significative e utili, appunto di valore, dai Big Data. Questo permette, nella migliore delle ipotesi, di poter estrarre vantaggio competitivo, migliorare l'efficienza operativa, innovare i processi aziendali e personalizzare l'esperienza del cliente. Per comprendere meglio il concetto di valore si riporta il celebre slogan coniato da Clive Humby, data scientist e matematico inglese nel 2006: "I dati sono il nuovo petrolio". Questo sta a testimoniare che così come il petrolio ha permesso lo sviluppo socioeconomico mondiale tra il XIX e XX secolo, a partire dal XXI i dati svolgeranno questo importante ruolo.

Si vuole ribadire quindi come i Big Data siano risorse informative ad alto volume, ad alta velocità e ad alta varietà che richiedono forme di elaborazione delle informazioni

⁴ Nasscom, Crisil launch 'Big Data-The Next Big Thing' report – 05/09/2012 - <https://www.moneycontrol.com/news/business/stocks/-1887295.html>

innovative ed economicamente sostenibili che favoriscono le intuizioni, supportano i processi decisionali e l'automazione dei processi.

1.6 Fonti

In accordo con una classificazione proposta dall'United Nations Economic Commissions for Europe (UNECE) esistono tre principali tipologie di fonti che possono essere considerate Big Data.

1. Dati generati dagli esseri umani:
 - a. Social Media e piattaforme online: Facebook, Twitter, Instagram, LinkedIn, blog, commenti, siti di recensioni ecc. ecc.: tutti generano una grande quantità di dati testuali, immagini e video che riflettono le opinioni, i comportamenti e le interazioni degli utenti.
 - b. Attività web: ricerche su internet tramite motori di ricerca (es. Google), video caricati su Internet (es. YouTube), prodotti visualizzati, pubblicità cliccate: questi rilasciano dati precisi sulle abitudini di consumo e sugli interessi degli utenti.
 - c. Dati geospaziali offrono informazioni sugli spostamenti e sulle preferenze di viaggio.
 - d. Altro: archivi di immagini e video, dati e contenuti provenienti da telefoni cellulari (es. messaggi di testo o livello di batteria) e da computer (es. modo di battere sulla tastiera), email ecc. ecc.
2. Dati generati dai processi:
 - a. Dati prodotti da enti e istituzioni pubbliche: es. cartelle cliniche
 - b. Dati prodotti dal settore privato: transazioni commerciali, registri di cassa, fatture, movimenti delle carte di credito ecc. ecc.
3. Dati generati da macchine
 - a. Dati provenienti da sensori fissi: domotica, varie tipologie di sensori: meteorologici, inquinamento, traffico, sicurezza, sorveglianza, salute ecc. ecc.

- b. Dati provenienti da sensori mobili, ovvero per scopi di tracciamento o analisi: es. localizzazione da telefoni cellulari o da dispositivi automobilistici.
- c. Dati provenienti da sistemi informatici.

1.7 Big Data Value Chain

Quando la dimensione dei dati diventa parte intrinseca del problema e costringe a superare le tecniche tradizionali prevalenti, diventa fondamentale saper padroneggiare nuove tecnologie. Questo è importante perché, come già descritto, i Big Data risultano essere uno dei principali asset economici del futuro. La capacità di sfruttarli in maniera efficace aumenterà la competitività delle imprese traducendosi in crescita economica e creazione di posti di lavoro. Per poterli saper gestire al meglio è necessario conoscere la Value Chain dei Big Data, rappresentata nella figura seguente:

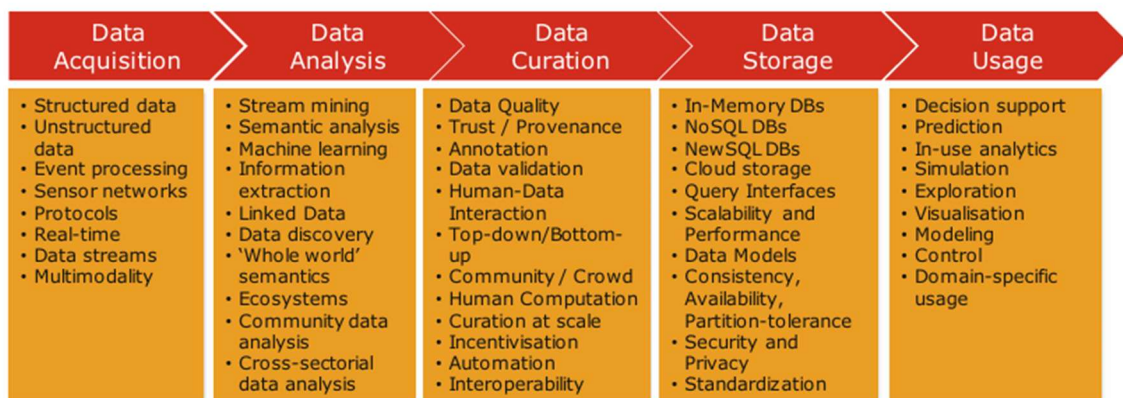


Figura 4: Data Value Chain - Data Economy

Descrivendo brevemente le fasi si ha:

1. L'acquisizione è il processo che consiste nel raccogliere, filtrare e pulire i dati prima di inserirli all'interno di un data warehouse o qualsiasi altra soluzione di archiviazione su cui è possibile effettuare l'analisi di dati.

Si possono acquisire dati strutturati o dati destrutturati. Questi possono essere presi da eventi in tempo reale (es. like su Facebook), da sensori distribuiti (es. IoT), da flussi continui che vanno gestiti in maniera costante (es. transazioni finanziarie). Infine, è importante sottolineare che per l'acquisizione dei dati esistono diversi protocolli, di cui la maggior parte non pubblici. Tra i più conosciuti si hanno: AMQP (Advanced, Message Queuing Protocol), Java Message Service, Storm, S4 e Kafka.

2. L'analisi dei Big Data consiste in una strutturazione dei dati grezzi (spesso presentati in formati differenti) che possa supportare il processo decisionale.

Per questa frase si utilizza un vecchio modo di dire inglese “turn lead into gold”, ovvero trasformare il piombo in oro proprio a voler sottolineare come l'amplessissima quantità di dati che può essere eterogenea per meccanismo di codifica, formato, struttura, semantica, provenienza, affidabilità e qualità sia trasformata in un insieme di dati utilizzabili e di importante valore.

Oggi il machine learning è una delle tecniche più utilizzate per l'analisi di questa grande quantità di dati. Questi algoritmi apprendono dai dati esistenti per identificare pattern, fare previsioni o classificazioni.

Un'altra tecnica è il data mining sui flussi che consiste di gestire flussi di dati ad alto volume e alta velocità che provengono da sensori o altre attività online in cui è coinvolto un alto numero di utenti.

Inoltre, per un'indicizzazione efficiente, estrazione e classificazione di entità e ricerca sui dati presenti sul web stanno prendendo sempre più piede i Linked Data ed approcci semantici. I primi permettono di collegare i dati provenienti da fonti diverse sul web andando a creare un grafo di informazioni accessibili e interpretabili dalle macchine, mentre il secondo consente di aggiungere significato a questi dati tramite ontologie, regole e linguaggi come OWL (Web Ontology Language). Entrambi trasformano i dati grezzi in conoscenza strutturata pronta per analisi avanzate.

Infine, è importante sottolineare l'importanza di un approccio cross-funzionale e comune che vada ad unire, integrare e condividere dati provenienti da più gruppi di utenti per ottenere insight più ampi e confrontabili.

3. La data Curation fornisce il supporto metodologico e tecnologico alla gestione dei dati per affrontare i problemi di qualità. Essa consiste nella gestione attiva e continua dei dati nel loro ciclo di vita di interesse e utilità.⁵ Questa fase permette la scoperta e il recupero dei dati, mantenendo buoni standard di qualità, aggiungendo del valore e permettendo il loro riutilizzo nel tempo. I processi della data curation possono essere categorizzati in diverse attività: creazione dei contenuti, selezione, classificazione, trasformazione, validazione e

⁵ Cragin et al, (2007)

conservazione. Essa permette la creazione di modelli basati sui dati più completi e di alta qualità. Tra le varie tecniche per realizzare questa fase si vuole citare: i Master Data Management (MDM) utili per rimuovere duplicati e standardizzare meglio i dati. Lo scopo principale è assicurare che l'organizzazione non usi dati incoerenti in diverse parti del sistema. I loro principali obiettivi sono:

- a. Sincronizzare i dati tra varie parti del sistema.
- b. Coordinare la gestione dei dati.
- c. Gestire la conformità e la presentazione dei dati in sistemi più analitici.

Un altro approccio utilizzato è la sheer curation o curation at source che permette di normalizzare e categorizzare meglio i dati.

Infine, anche qui è importante un approccio di tipo collaborativo. Con il crowdsourcing, infatti, si coinvolgono grandi gruppi di persone per condividere costi e compiti complessi.

4. L'archiviazione dei Big Data riguarda la conservazione e gestione dei dati in modo scalabile soddisfacendo le varie richieste di accesso ai dati da parte delle applicazioni. Il sistema ideale è quello che permetterebbe di conservare una quantità praticamente illimitata di dati, affrontare alti tassi di accessi casuali sia in lettura che scrittura, supportare sia dati strutturati sia non strutturati, essere flessibile così da accogliere una vasta tipologia di dati e infine, per motivi di privacy, dovrebbe lavorare solamente su dati criptati. È altrettanto importante che il database non vada a rallentare notevolmente la soddisfazione della query richiesta. Ad oggi sono presenti diversi sistemi di archiviazione:

- a. Distributed File Systems, ad esempio Hadoop File System (HDFS) che offre la possibilità di archiviare grandi quantità di dati non strutturati in modo affidabile su hardware commerciale.
- b. NoSQL Databases: è la famiglia di tecnologie di archiviazione per Big Data più importante. Non si basano su modelli relazionali e sono pensati per offrire una maggiore scalabilità a scapito di una minore consistenza.
- c. NewSQL Databases: è una forma moderna di database che vuole garantire una scalabilità comparabile alla categoria precedente mantenendo le garanzie transazionali fornite dai sistemi di database tradizionali: migliore atomicità, consistenza, isolamento e durabilità.

- d. Big Data Querying Platforms: il loro obiettivo è fornire un'interfaccia di alto livello e ridurre al minimo i tempi di risposta
5. Data Usage: come già anticipato è fondamentale che i dati siano utilizzati a supporto delle decisioni organizzative. Altri principali ambiti di utilizzo riguardano l'analisi predittiva, la simulazione e la modellizzazione. È importante evidenziare le principali aree dei casi d'uso dei Big Data siano: il settore della produzione, dei trasporti, della logistica, oltre che il marketing e la profilazione avanzata.

1.8 Tecnologie abilitanti

Dai paragrafi che precedono risulta evidente come l'infrastruttura tecnologica svolga un ruolo di fondamentale importanza per il processamento dei dati. In questo paragrafo si vogliono citare le più determinanti:

- Internet a banda larga: dagli anni 2000 è passato progressivamente dal 3G al 5G rendendo possibile la trasmissione di enormi volumi di dati con una latenza minima.

A tal proposito si riportano i seguenti dati⁶:

Velocità di download dati, picco:

- 3G: dai 2 Mbit/s ai 42 Mbit/s.
- 5G: 10 Gbit/s.

Andando a fare un confronto sul massimo dei valori si nota un incremento di 3 ordini di grandezza.

Velocità di download dati media:

- 3G: dai 384 kbit/s agli 8 Mbit/s.
- 5G: dai 150 Mbit/s ai 200 Mbit/s.

⁶ Mobile Data Speed with 2G, 3G, 4G and 5G Cellular Network – Adnan Ghayas – 07/01/2020 – COMMSBRIEF – <https://commsbrief.com/mobile-data-speed-with-2g-3g-4g-and-5g-cellular-networks/>
Timeline from 1G to 5G: A Brief History on Cell Phones – Richard Galazzo – 21/09/2020 – CENG – <https://www.ceng.ca/information-centre/innovation/timeline-from-1g-to-5g-a-brief-history-on-cell-phones/>

Questo salto ha permesso di migliorare significativamente l'esperienza del cliente, ad esempio, nel commercio al dettaglio, riducendo i tempi d'attesa sia negli acquisti online sia nei punti vendita fisici. Questi servizi digitali infatti offrono significativi risparmi di tempo, aumentano l'utilità e contribuiscono alla crescita del reddito familiare.

Importante sottolineare che la banda larga favorisce la fruizione di contenuti multimediali complessi ed interattivi (Es. Facebook, Google Search, Instagram, Wikipedia...) ma abilita anche per esempio tutto ciò che è l'internet of Things.

Secondo dati del McKinsey Global Institute le aziende basate sui dati con connessione 5G hanno 23 volte più probabilità di acquisire clienti, 19 volte più probabilità di essere redditizie e 6 volte più probabilità di fidelizzare dei buoni clienti. Questo testimonia come l'infrastruttura di rete sia una tecnologia abilitante fondamentale per l'economia dei Big Data e per un miglioramento della competitività d'azienda.

Infine, è importante sottolineare che il consumatore ha accesso a servizi finanziari, sanitari, pubblici e all'e-commerce in modo più rapido. Questi lo portano a ottenere un maggior benessere e dunque ad una più grande disponibilità a pagare per connessioni migliori.

- Internet of Things: in termini di dispositivi gli smartphone contribuiscono in maniera significativa agli odierni sistemi di tracciamento e profilazione. Questi, infatti, forniscono dettagli sulla personalità e sulla vita di tutti i giorni dell'utente. Nei recenti anni, molti altri devices dotati di sensori e connessioni di rete sono entrati nella vita dei consumatori aggiungendo modi di tracciamento. Questi includono: e-book, dispositivi indossabili, smart TV, contatori, termostati, rilevatori di fumo, frigoriferi, occhiali, spazzolini da denti e giocattoli. Tutti questi offrono alle aziende un accesso senza precedenti ai comportamenti dei consumatori in un'ampia varietà di momenti della vita.

Un altro esempio riguarda i dati generati dai veicoli: le auto sempre più ricche di sensori, riescono a fornire moltissime informazioni di interesse non solo per i produttori ma anche per concessionari, compagnie assicurative, istituti di credito, fornitori di servizi telematici ecc. ecc.

➤ Machine learning: è lo studio e l'applicazione di algoritmi che permettono ai programmi informatici di migliorare automaticamente attraverso l'esperienza, adattandosi e apprendendo autonomamente da nuovi dati per identificare schemi e ricavare informazioni. Gli algoritmi di machine learning elaborano grandi quantità di dati per apprendere strategie e comportamenti da adottare in contesti specifici. Il machine learning, dunque, si serve dei Big Data per poter estrarre valore, ottimizzare i processi e supportare le decisioni.

➤ Cloud computing: si tratta di un insieme di server ad alte prestazioni in grado di immagazzinare, visualizzare e interrogare grandi insiemi di dati molto più velocemente rispetto a un computer standard. Questa struttura, basandosi su server remoti multipli, può ricevere enormi "raffiche" di dati da sistemi ad alta intensità ed interpretarli in tempo reale. In generale, il cloud computing completa i Big Data e fornisce un ambiente di calcolo comodo, su richiesta e condiviso, con uno sforzo minimo di gestione e ridotta complessità. Inoltre, offre un ambiente più robusto, più automatizzato e facilmente scalabile.

È importante sottolineare come sia necessario gestire le vulnerabilità di sicurezza: nell'integrazione tra cloud e Big Data, è necessario implementare politiche di sicurezza e piani di consolidamento adottando strumenti quali autenticazione, controllo degli accessi, crittografia, registrazione e monitoraggio degli eventi.

In conclusione, il cloud assicura, grazie a una grande disponibilità di archiviazione e una forte potenza di elaborazione, uno spazio di lavoro eccellente per tutte le esigenze dei Big Data.

➤ Servizi digitali e applicazioni online: la crescente diffusione di piattaforme digitali, e-commerce, social media ed app mobile consente una raccolta dettagliata e sistemica di un'enorme quantità di informazioni sui consumatori. Questi Big Data riguardano preferenze ed abitudini d'acquisto, comportamenti di navigazione, rete di contatti, oltre ad analisi psicologiche e comportamentali. L'analisi di questi dati permette alle aziende o organizzazioni di generare insight strategici e migliori strategie di marketing.

➤ Cookies: sono creati dal server e costituiscono dei mezzi per raccogliere informazioni generate da un sito web e salvate da un browser di un utente internet

sul disco rigido locale del computer o del dispositivo mobile. Essi svolgono la funzione di semplificare e velocizzare gli accessi ai siti web da parte degli utenti, sono inviati direttamente al browser dell'utente e salvati direttamente nel computer dal quale si accede alla sessione. Il compito di molti cookies è tenere traccia delle abitudini degli utenti del web. Essi sono usati per monitorare e profilare gli utenti durante la navigazione, studiano i loro movimenti e le loro abitudini di consultazione del web o di consumo ai fini del behavioural advertising.

Essi si distinguono in:

- a. Cookies di prima parte: creati dal web server del sito che è visitato dall'utente. Sono generati per personalizzare l'esperienza di navigazione riconoscendo automaticamente un utente che è già entrato su quel sito. Essi sono leggibili solamente sul dominio su cui sono stati creati e non all'esterno. Risultano quindi inefficaci per quanto riguarda l'erogazione di pubblicità personalizzata su altri siti.
- b. Cookies di terza parte: creati da domini diversi da quelli del sito che l'utente sta visitando. Sono utilizzati per il tracciamento degli utenti al fine di fornire annunci pubblicitari rilevanti e personalizzati.
- c. Cookies di seconda parte: tipologia di cookies che è trasferita da un'azienda che li ha creati come cookies di prima parte ad un'altra società, attraverso ad esempio una partnership.

Dall'analisi delle tecnologie abilitanti si evince come l'infrastruttura digitale rappresenti un elemento fondamentale per l'economia dei Big Data. La sinergia tra queste tecnologie consente alle imprese di generare insight strategici, ottimizzare processi decisionali e sviluppare strategie di marketing avanzate. È di fondamentale importanza però, un'attenzione costante agli aspetti legati alla sicurezza dei dati e alla tutela della privacy.

1.9 Tipologie di dati

Si riportano infine, per completezza, varie tipologie di classificazioni dei dati.

La prima distingue i dati in strutturati, destrutturati e semi-strutturati

- Strutturati: hanno uno schema fisso e si inseriscono ad esempio ordinatamente in righe e colonne. Questo, prende il nome di schema on-write. Uno dei più classici

esempi di dati strutturati è il database relazionale: qui i dati sono formattati in campi definiti con precisione, ad esempio: numero di carta di credito, codice fiscale oppure codice ID.

Esempi comuni di dati strutturati sono quelli generati dai punti vendita tramite la lettura dei codici a barre.

- **Destrutturati:** non hanno uno schema fisso e possono avere formati più complessi. Sono archiviati nel loro formato nativo (schema on read) e non sono elaborati sino a quando non sono utilizzati. Si presentano in un'infinità di formati di file diversi, tra cui email, post sui social media, presentazioni, chat, dati dei sensori IoT ecc. ecc. si prevede che alla fine del 2025 l'80% dei dati sia destrutturato.⁷
- **Semi-strutturati:** combinazione di dati strutturati e non strutturati. Qui i dati contengono informazioni sufficienti da consentire la loro catalogazione.

Una seconda classificazione che si può andare a realizzare è tra dati esterni e dati interni:

- I primi sono quelli che non provengono direttamente dall'organizzazione stessa ma appunto dall'esterno. Essi non sono direttamente controllati ma sono utili per arricchire i processi decisionali.
- I secondi invece sono generati direttamente dall'organizzazione tramite i sistemi, i processi e gli utenti. Questi spesso possono contenere informazioni strategiche o sensibili.

Un ulteriore criterio di classificazione fa riferimento al momento di raccolta e all'utilizzo dei dati.

- **Batch data:** qui la computazione e l'analisi dei dati avviene in un momento successivo rispetto alla loro raccolta, sono effettuate su un insieme più grande di dati.
- **Real time data:** qui l'elaborazione è istantanea e continua, non c'è quasi differenza temporale con l'avvenimento dell'evento.

⁷ Big Data & Advanced Analytics per il Risk Management – Associazione Italian Financial Industry Risk Managers.

Inoltre, i dati possono essere distinti tra dati reali e sintetici. Questi ultimi fanno riferimento a quei dati creati artificialmente dalle macchine imitando quelli che potrebbero essere ottenuti dal mondo reale. Essi sono prodotti utilizzando algoritmi di machine learning di tipo generativo. Spesso sono utilizzati per il training di algoritmi e per la simulazione di situazioni nuove o complesse raramente incontrabili.

Infine, si possono distinguere i:

- Dati Transazionali: dati raccolti delle transazioni che possono essere classificate come finanziarie, operative, digitali o logistiche. Il loro scopo è quello di supportare l'organizzazione nelle operazioni quotidiane.
- Dati analitici: sono originati dai dati transazionali, infatti rappresentano una loro riorganizzazione ed elaborazione in modo da rendere chiaramente visibili le loro informazioni veramente utili.
- Dati Translytics: dati transazionali che sono analizzati in tempo reale fornendo così insights e previsioni pressoché istantanee.

Al termine di questo paragrafo si vuole fornire un'ultima descrizione sintetica di quelle che sono le tipologie di dati raccolti e analizzati dai consumatori.

Innanzitutto, si parla di dati di prima parte quando i consumatori instaurano una relazione diretta con l'organizzazione che ha raccolto gli stessi dati mentre di dati di terza parte quando i dati sono elaborati per conto di altre aziende oppure raccolti, acquistati o concessi in licenza da altri.

Infine, riferendosi ai dati personali, essi sono raggruppati in 4 categorie differenti:

- Volontari: dati esplicitamente condivisi da parte degli individui.
- Osservati: acquisiti registrando le attività dei consumatori.
- Effettivi: si riferiscono a informazioni fattuali sugli individui.
- Modellati: derivano da inferenze sulle caratteristiche o sui comportamenti previsti dei consumatori basandosi su dati effettivi.

2 – IL MERCATO DEI BIG DATA

2.1 Introduzione al mercato

Cent'anni fa, nell'arco della vita della persona, solo pochi dati erano raccolti: la nascita, l'eventuale matrimonio e la morte. Questi dettagli, ironia a parte, erano così brevi ed essenziali da poter essere riassunti su una lapide. L'idea dei mercati dei dati personali risale a Laudon⁸ nel 1996, che immaginava la creazione di un mercato nazionale dell'informazione, in cui gli individui avrebbero potuto depositare le proprie informazioni in edifici istituzionali simili a banche per poi ottenere un compenso derivante dall'utilizzo dei propri dati.

Oggi, la situazione è decisamente diversa: ogni volta che si utilizza uno smartphone, si naviga su una connessione a internet, si striscia una carta di credito, si acquista un oggetto o si prende un trasporto pubblico con biglietto digitale, sono lasciate, deliberatamente o meno, tracce del nostro passaggio, delle nostre preferenze e delle nostre scelte. A questi fenomeni contribuiscono differenti fattori:

1. Crescente digitalizzazione che porta alla produzione di dati sia strutturati che non strutturati che diventano sempre più importanti per i processi decisionali aziendali e per l'efficienza operativa.
2. Uso di internet in mobilità: dispositivi digitali e IoT che generano quotidianamente volumi crescenti di dati.
3. Automazione e robotizzazione: le procedure automatizzate producono e accumulano informazioni che possono rivelarsi molto preziose.
4. Personalizzazione: permette di offrire servizi e prodotti sempre più su misura che migliorano l'esperienza dell'utente/cliente rendendola unica e coinvolgente.
5. Disponibilità crescente di dati pubblici: queste informazioni diventano parte integrante delle analisi dei Big Data.
6. Intelligenza artificiale e machine learning: il loro sviluppo permette un miglior sfruttamento dei dati raccolti con la creazione di oggetti/servizi intelligenti e personalizzati.

⁸ Kenneth C. Laudon, professore di Information Systems alla New York University (Stern School of Business) che nel 1996 ha pubblicato il saggio "Markets and Privacy"

Il mercato dei Big Data sta vivendo una crescita senza precedenti anche perché le organizzazioni riconoscono sempre più il potere trasformativo derivante dallo sfruttamento di enormi quantità di dati. Esse, infatti, attuano processi decisionali più strategici, migliori esperienze per i clienti e una maggiore efficienza operativa. Inoltre, grazie alla diffusione della trasformazione digitale, la domanda di analisi avanzata e di elaborazione di informazioni in tempo reale contribuisce in modo significativo alla crescita di questo mercato.

Quanto appena descritto porta l'Economist a stimare una crescita della produzione dei dati digitali di circa il 45% annuo.

Viste tutte le ragioni evidenziate qui sopra, i dati sono stati definiti, come anticipato nel capitolo precedente, il petrolio dell'era digitale. La metafora è ben rappresentata dall'immagine seguente, pubblicata sulla rivista The Economist⁹ del 06/05/2017. Infatti, si vedono diverse piattaforme digitali e grandi aziende tecnologiche come Amazon, Uber, Google, Microsoft, Facebook e Tesla che sono raffigurate come piattaforme petrolifere in mezzo al mare. Queste grandi aziende riescono ad estrarre enormi quantità di informazioni dagli utenti (es. comportamenti, preferenze, spostamenti e consumi). I dati una volta raccolti, raffinati e analizzati diventano la base per creare un vantaggio competitivo da parte dell'azienda. Un altro lato della metafora, che è opportuno sottolineare, è legato ai rischi che si possono formare: così come il petrolio ha avuto e continua ad avere effetti collaterali ambientali, sociali e geopolitici gli stessi dati sollevano numerosissime questioni etiche, giuridiche e politiche.



Figura 5: The world's most valuable resource - The Economist

⁹ The world's most valuable resource is no longer oil but data – 05/05/2017 – The Economist – <https://www.economist.com/leaders/2017/05/06/the-worlds-most-valuable-resource-is-no-longer-oil-but-data>

2.2 Le dimensioni del mercato

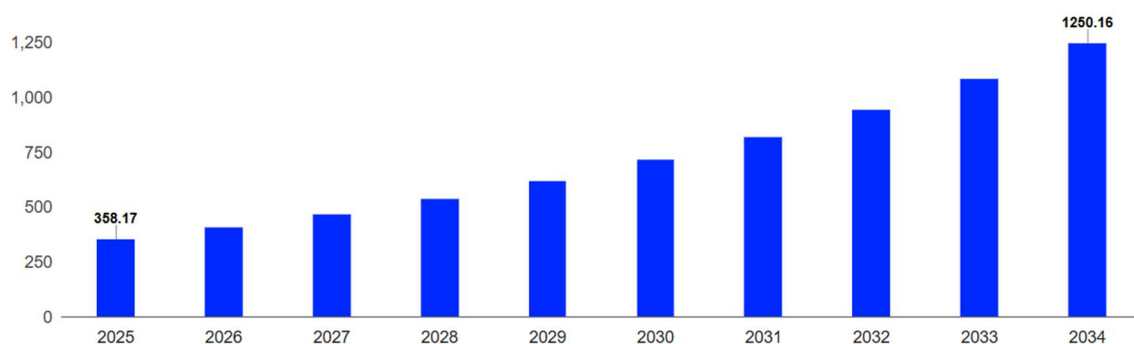
A seguito delle ricerche effettuate, si è osservato che le stime sul valore e sulla crescita del mercato dei Big Data presentano una certa variabilità a seconda delle fonti considerate. Le differenze dipendono sia dall'orizzonte temporale: le stime più brevi (circa 5 anni) tendono a essere più prudenti rispetto a quelle su periodi più lunghi (10-15 anni), sia dal perimetro del mercato preso in considerazione: alcune si riferiscono esclusivamente al mercato dei Big Data altre invece includono anche la parte di Business Analytics.

Al di là delle differenze numeriche, la tendenza generale risulta decisamente chiara: il mercato dei Big Data è destinato a una crescita significativa e a giocare un ruolo sempre più centrale nei processi decisionali, nell'efficienza operativa e nella creazione di vantaggio competitivo per le imprese.

Di seguito, si illustrano i principali valori riscontrati, a conferma di quanto sopra descritto.

Secondo Expert Market Research (EMR)¹⁰, una società di ricerca di mercato e consulenza aziendale nota anche come Claight Corporation, valuta il mercato dei Big Data, considerando anche la Business Analytics, a 311,72 miliardi di dollari nel 2024 con una crescita attesa sino a 1250 miliardi di dollari nel 2035, con un tasso di crescita composto annuo (CAGR) pari al 14.90% tra il 2025 e il 2035.

Nel grafico sottostante si riporta un possibile andamento del mercato dei Big Data secondo l'analisi appena presentata:



*Figura 6: Big Data Market Size and Share Outlook, Forecast Trends and Growth Analysis Report - EMR
CLAIGHT*

¹⁰ Big Data Market Size and Share Outlook – Forecast Trends and Growth Analysis Report – EMR
CLAIGHT – <https://www.expertmarketresearch.com/reports/big-data-market>

Secondo MarketsandMarkets^{TM11}, inclusa da Forbes nella lista delle migliori società di consulenza gestionale negli Stati Uniti del 2025, stima che il solo mercato dei Big Data abbia raggiunto i 220,2 miliardi di dollari nel 2023 e che raggiungerà nel 2028 i 401,2 miliardi di dollari a un tasso di crescita composto annuo del 12.7%.

Infine, secondo Market Research Future¹² sostiene che il solo mercato dei Big Data possa raggiungere i 250 miliardi di dollari nel 2035 ad un tasso di crescita annuo composto del 10.58%.

2.2.1 In Italia

In Italia, secondo l'Osservatorio Big Data & Analytics del Politecnico di Milano il mercato dei Big Data ha raggiunto, nel 2024, un valore di 3,42 miliardi di euro con una crescita media annua dal 2021 ad oggi del 20%. Andamento che sembra essere fortemente trainato anche dal crescente interesse per la Generative AI.

Secondo Carlo Vercelli, responsabile dell'Osservatorio, le aziende italiane attualmente stanno progressivamente costruendo una propria strategia di valore basata sull'utilizzo diffuso e consapevole dei dati. Questo si basa su tre principi fondamentali:

1. Solida gestione e usabilità dei dati.
2. Capacità di innovazione e di implementazione di progettualità di analisi avanzata.
3. Evoluzione culturale orientata alla diffusione della Data & AI Literacy.

La spesa nel mercato dei Big Data è dominata da parte delle grandi imprese, che rappresentano il 75% del totale, seguite dalle PMI e le microimprese con il 19% e infine dalla Pubblica Amministrazione con il 6%.

L'analisi evidenzia inoltre che nel 2024 circa il 79% delle PMI (5 punti percentuali in più rispetto al 2023) svolge analisi di dati almeno a livello predittivo. Tuttavia, le compiono in maniera sporadica e senza utilizzare figure professionali dedicate. Ne consegue che la

¹¹ Big Data Market – MARKETSANDMARKETSTM – Gennaio 2024 –

<https://www.marketsandmarkets.com/Market-Reports/big-data-market-1068.html>

¹² Big Data Market Summary – Market Research Future –

<https://www.marketresearchfuture.com/reports/big-data-market-7846>

formazione del personale e l'upskilling rimangano centrali per colmare il gap di competenze e avvicinare l'Italia alla media mondiale in termini di maturità dei processi di valorizzazione dei dati.

Il grafico seguente, realizzato dal Politecnico di Milano, riporta l'andamento del mercato dei Big Data italiano. A tal proposito si vuole sottolineare infine, come le considerazioni effettuate a livello globale nella sezione precedente, sembrano riflettersi in maniera significativa nel contesto nazionale.

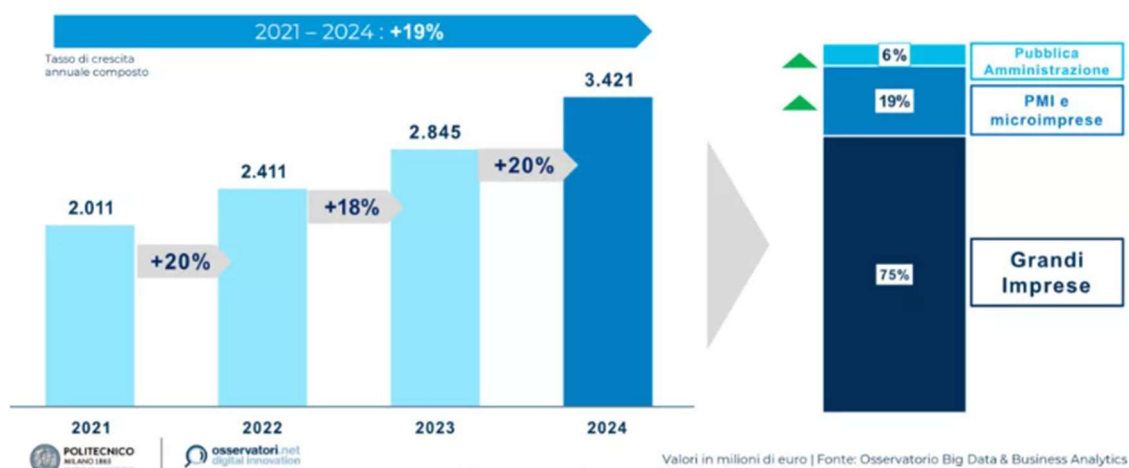


Figura 7: Il mercato italiano dei Big Data - POLIMI School of Management

2.3 Gli attori in generale

In questo paragrafo si intende mettere in evidenza le figure che partecipano a questo mercato, senza citare persone o aziende specifiche, ma descrivendo in astratto i soggetti coinvolti.

2.3.1 Consumatori

I consumatori sono coloro che creano e rivelano i propri dati personali come sottoprodotto dell'utilizzo dei servizi offerti dai fornitori di contenuti digitali (CSP), ad esempio tramite la navigazione sul web, acquisti online o l'uso dei social media.

Ad essi è riconosciuto, tramite l'articolo 20 del Regolamento Generale sulla Protezione dei Dati (GDPR) dell'Unione Europea e la Sezione 1798.130(a) del California Consumer Privacy Act (CCPA), il diritto alla portabilità dei dati: essi possono accedere gratuitamente alle proprie informazioni e decidere liberamente a chi cederle. A tal

proposito possono decidere di concedere l'accesso ai propri dati ai Personal Data Brokers (PDB) qualora ricevano da questi un'adeguata ricompensa.

Per facilitare questo processo, i consumatori possono ricorrere ai Personal Information Management Systems (PIMS) che consentono di collegare e arricchire i dati generati durante l'uso dei CSP e di condividerli con terze parti tramite meccanismi di controllo degli accessi.

2.3.2 Content and Service Providers – CSP

I Content and Service Providers (CSP) sono le grandi piattaforme e i fornitori di servizi digitali che basano i propri modelli di business principalmente sulla raccolta dei dati degli utenti e sulla loro monetizzazione in vari modi, ad esempio attraverso la pubblicità mirata o analisi dei comportamenti dei consumatori.

Il servizio offerto è spesso gratuito o a basso costo, infatti, il vero valore deriva dalla raccolta e successiva monetizzazione dei dati generati dagli utenti. Questi ultimi non ricevono alcun compenso diretto ma possono usufruire delle funzioni messe a disposizione dal servizio.

Nel modello che si andrà ad analizzare i CSP reagiscono strategicamente all'ingresso dei Personal Data Brokers (PDB), adattando la qualità e il prezzo dei propri servizi. Essi dispongono comunque di un vantaggio competitivo strutturale in quanto già in possesso di grandi quantità di dati derivanti dall'erogazione dei propri servizi digitali.

2.3.3 Personal Information Management Systems – PIMS

I Personal Information Management Systems (PIMS) sono sistemi che permettono agli utenti e ai consumatori di conservare, gestire e monitorare i propri dati provenienti dai vari CSP, mantenendone il pieno controllo.

Questi sistemi permettono, ad esempio, di aggregare in un unico ambiente dati da più fonti: finanziari personali (come informazioni su carte di credito), cronologie di acquisti e contenuti generati sui social media.

2.3.4 Personal Data Brokers – PDB

I Personal Data Broker (PDB) possono essere assimilati alle banche di dati personali immaginate da Laudon nel 1996. Essi, infatti, consentono ai consumatori di monetizzare

i propri dati personali, raccogliendoli, aggregandoli e gestendoli su mandato e consenso degli utenti. In cambio, i consumatori ricevono una ricompensa dai PDB, i quali successivamente si occupano della commercializzazione dei dati raccolti. In questo senso si evince che i PDB operano in concorrenza con i grandi CSP.

Grazie al diritto alla portabilità dei dati concesso ai consumatori i PDB non sono obbligati a negoziare con i CSP per ottenere l'accesso alle informazioni ma possono acquisirle direttamente dai consumatori. Una volta effettuato il trasferimento, il PDB possiede una copia dei dati provenienti dai CSP e in qualità di titolare può decidere, nel rispetto del consenso dell'utente, come monetizzarli.

In molti casi, i PDB integrano direttamente funzionalità tipiche dei PIMS come la possibilità di aggregare dati provenienti da fonti diverse e facilitare lo scambio di informazioni.

Il ruolo teorico dei PDB sarebbe quello di redistribuire il valore dei dati verso i consumatori, ponendosi come intermediari tra utenti e mercato. Con la loro presenza si vorrebbe offrire un modello alternativo rispetto al predominio, quasi impossibile da gestire e controllare, dei CSP. Tuttavia, diversi studi mostrano che tali attori incontrano notevoli difficoltà a raggiungere una massa critica di fornitori di dati nonché a garantire ricompense realmente significative. Inoltre, il benessere degli utenti finali, se sono presenti i PDB sembra diminuire.

Di conseguenza possiamo affermare che la sostenibilità dei PDB come attori di mercato appare piuttosto limitata; queste considerazioni saranno analizzate attentamente successivamente.

2.3.5 Regolatori e policy makers

I regolatori e i policy makers giocano un ruolo fondamentale nel definire le regole e i meccanismi che governano il mercato dei dati, incidendo soprattutto sulla tutela dei consumatori. Le principali autorità di regolamentazione saranno citate nel paragrafo successivo, inoltre, nel Capitolo 4 sarà analizzato un caso studio in cui esse hanno svolto un ruolo di primaria importanza.

2.3.6 Esempi pratici

Una volta presentati in maniera astratta i vari attori del mercato dei Big Data, si vuole descrivere in maniera estremamente sintetica ciascuno di essi in termini pratici.

Tra i consumatori si possono considerare, ad esempio, gli utenti di Facebook o i clienti che si registrano e compiono operazioni su Amazon.

Tra i Content and Service Providers rientrano piattaforme come Facebook, Google, Amazon, che utilizzano i dati degli utenti per migliorare i servizi offerti e per scopi pubblicitari.

Tra i Personal Information Management Systems si trova ad esempio la piattaforma Digi.me che permette agli utenti di aggregare i propri dati da diverse fonti, permettendo loro il controllo su come e con chi dividerli e offrendo una maggiore privacy nella gestione delle informazioni personali.

Tra i principali Personal Data Brokers si hanno aziende come Acxiom, Experian ed Equifax, che raccolgono e commercializzano dati personali a terzi.

Infine, tra i principali regolatori si evidenziano: la Commissione Europea che ha introdotto il Regolamento Generale sulla Protezione dei Dati (GDPR), il California Consumer Privacy Act (CCPA), la Federal Trade Commission (FTC) negli USA, l'European Data Protection Supervisor (EDPS) e, in Italia, il Garante per la protezione dei dati personali.

2.4 Approfondimento sui Personal Data Brokers

Come anticipato nella sezione precedente si intende evidenziare come, sulla base di un modello della teoria dei giochi, elaborato da Jan Krämer, Daniel Schnurr e Bastian Haberer nell'articolo *Do Consumers Benefit from Selling Their Data? The Economic Impact of Personal Data Brokers on Digital Markets*, la figura del Personal Data Broker difficilmente riuscirà ad affermarsi in maniera solida nei mercati digitali, portando all'impossibilità di stabilire mercati dei dati personali efficienti e duraturi. Inoltre, dagli studi emerge come, in quasi tutti i casi, tale attore vada a danneggiare persino il benessere totale dei consumatori.

Il modello analizzato presenta due principali esiti di mercato che variano a seconda dell'efficienza del PDB nel saper monetizzare i dati.

- Se l'efficienza del PDB è bassa, ovvero offre ai consumatori una ricompensa minima, appena sufficiente da garantire la loro partecipazione, lo scenario prende il nome di equilibrio di reddito minimo. In questo scenario, a seguito della concorrenza creata dal PDB, il CSP sarà invogliato a peggiorare la qualità del proprio servizio. Il mercato dei dati personali fallisce e il surplus dei consumatori risulta inferiore rispetto a uno scenario privo di PDB.
- Se il PDB è altamente efficiente, ovvero offre ai consumatori ricompense significative, questo scenario prende il nome di equilibrio di reddito positivo. Tuttavia, in tale situazione i consumatori possono ancora trovarsi in una condizione peggiore rispetto ad un mercato senza PDB: il CSP, infatti, osservando che il PDB offre una ricompensa elevata e quindi spinge il consumatore ad utilizzare il servizio indipendentemente dalle sue caratteristiche, può decidere di abbassare ulteriormente la qualità e per di più alzarne il prezzo sapendo che i consumatori, imperterriti, continueranno ad utilizzarlo e dunque a generare dati grazie agli incentivi offerti dal PDB.

Nella pratica molti PDB sono attualmente bloccati nell'equilibrio di reddito minimo, poiché le ricompense offerte agli utenti sono spesso molto basse, nonostante negli ultimi anni l'ente normativo abbia fatto aumentare la consapevolezza tra i consumatori del valore economico dei dati personali.

Il modello proposto permette di analizzare i trade-off strategici che emergono con l'ingresso dei PDB nel mercato e di identificare le condizioni economiche sotto le quali i PDB possono prosperare nei mercati dei dati personali. Tuttavia, mostra anche che, laddove questi mercati riescano a consolidarsi, i consumatori rischiano comunque di trovarsi in una condizione complessivamente peggiore.

Prima di iniziare a descrivere i modelli è bene andare a sottolineare la tempistica del gioco, rappresentata nella figura sottostante:

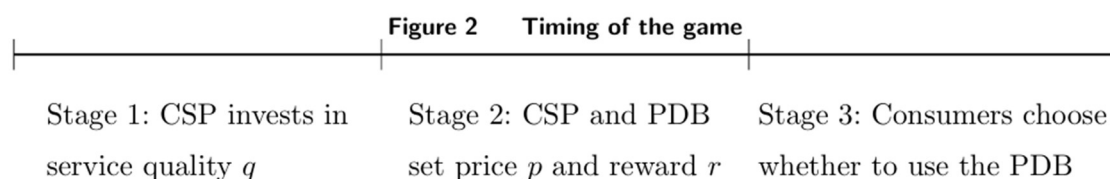


Figura 8: stages del gioco - The Economic Impact of Personal Data Brokers

1. Nel primo stadio (scelta di lungo termine) il CSP investe sulla qualità del servizio da erogare.
2. Nel secondo stadio il CSP fissa il prezzo e il PDB fissa la ricompensa per i consumatori.
3. Nel terzo stadio i consumatori scelgono se utilizzare o meno il PDB.

2.4.1 Modello base senza PDB

Without Personal Data Broker (M)

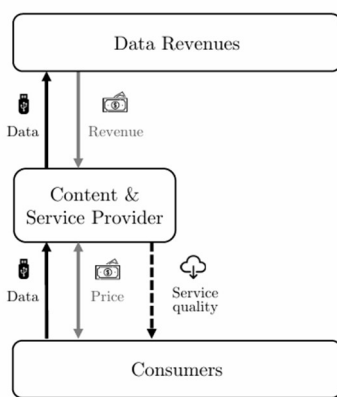


Figura 9: Modello con senza PDB - The Economic Impact of Personal Data

All'interno di questo modello il CSP ha due flussi di ricavi: in primo luogo, può addebitare ai consumatori un prezzo (che può essere anche negativo o pari a zero) per l'utilizzo del suo servizio. In secondo luogo, può generare dei ricavi andando a monetizzare sui dati dei consumatori.

Il CSP va a determinare strategicamente sia la qualità del proprio servizio sia il prezzo d'uso andando a massimizzare il proprio profitto totale.

Si nota immediatamente che: quanto più alta è la qualità del servizio o più basso il prezzo, tanti consumatori lo utilizzeranno e di conseguenza andranno a fornire più dati al CSP.

2.4.2 Modello base con PDB

In questo modello i consumatori possono decidere di trasferire i propri dati al PDB in cambio di una ricompensa monetaria da parte di quest'ultimo. Il PDB monetizza tali dati in concorrenza con il CSP per i ricavi derivanti dalle informazioni.

Per le restanti componenti la struttura del modello non cambia.

With Personal Data Broker (MP)

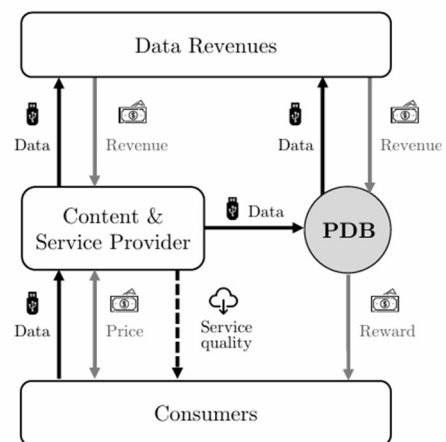


Figura 10: Modello con PDB - The Economic Impact of Personal Data Brokers

2.4.3 Assunzioni del modello base

1. La domanda del consumatore per il servizio fornito dal CSP è limitata e il consumatore sceglie il consumo x che massimizza la propria utilità.
2. I costi marginali per la fornitura del servizio sono pari a 0.
3. Fornire dati al PDB non comporta costi di transazione; dunque, il consumatore utilizza il PDB se la ricompensa è positiva. Un esempio di costi di transazione può essere il costo di trasferimento dati.
4. I consumatori hanno il diritto e i mezzi per trasferire i propri dati al PDB.

2.4.3 Considerazioni sui consumatori

Nel modello si considera una massa unitaria di consumatori con una domanda totale per il servizio del CSP pari $x(q, p, r)$ in cui:

- q indica la qualità del servizio
- p indica il prezzo del servizio
- r indica la ricompensa pagata dal PDB, nel caso in cui esista.

È chiaro che il consumatore, durante l'utilizzo del servizio (es. digitando parole chiave, lasciando Mi Piace o effettuando transazioni), rilascia dati personali che rivelano le proprie preferenze. Per semplicità del modello si assume che la quantità dei dati raccolta dal CSP sia direttamente proporzionale al livello di utilizzo dei consumatori e per comodità sono indicati entrambi con la lettera x .

Qualora il consumatore decidesse di usufruire del PDB, trasferirebbe a quest'ultimo i propri dati ricevendo una ricompensa pari a $R(x) = r * x$.

Si assume che l'utilizzo del servizio del CSP da parte dei consumatori, x , aumenti all'aumentare della qualità del servizio q , della ricompensa del PDB r ma diminuisca all'aumentare del prezzo p scelto dal CSP.

Di conseguenza si assume come funzione di domanda del modello pari a:

$$x(q, p, r) = q - p + r$$

Questa funzione di domanda si ottiene andando a costruire una funzione utilità del consumatore:

- Assumendo un'interazione positiva tra qualità e utilizzo del servizio si può affermare che l'utilità del consumatore cresce all'aumentare dell'utilizzo del servizio fornito dal CSP.

Indichiamo questo termine con $Q(x) = q * x$

- È importante sottolineare che, se il consumatore passa del proprio tempo sul servizio proposto dal CSP non lo trascorre altrove; dunque, si troverà ad avere meno tempo a disposizione per svolgere altre attività (il tempo del consumatore è ovviamente limitato). Riferendosi ai costi opportunità, l'utilizzo del servizio diventa sempre più costoso, poiché interferisce sempre di più con altre attività preziose per il consumatore. Si indica questo termine con la funzione di costo convessa: $\theta(x) = \frac{1}{2}x^2$.
- Il consumatore per poter utilizzare il servizio dovrà sostenere un prezzo p che può essere positivo, pari a 0 oppure negativo (dunque, per l'assunzione 2, sotto il costo marginale). Si indica questo termine con $P(x) = p * x$.
- Infine, se i consumatori decidono di utilizzare il PDB, riceveranno una remunerazione che sarà proporzionale all'utilizzo del servizio (più lo utilizzano, più dati riescono a elaborare e dunque più dati trasferiscono). Indichiamo questo termine con: $R(x) = r * x$.

Definiti i quattro termini si può scrivere la funzione di utilità del consumatore come:

$$U = Q(x) - \theta(x) - P(x) + R(x)$$

Sostituendo con quanto descritto, si ottiene:

$$U = q * x - \frac{1}{2}x^2 - p * x + r * x$$

Massimizzando l'utilità del consumatore, e quindi applicando la condizione del primo ordine si ottiene:

$$\frac{dU}{dx} = 0 \rightarrow q - x - p + r = 0 \rightarrow x = q - p + r$$

In una versione più completa e formale si scrive:

$$x^*(p, r, q) = \arg \max [U(x, p, r, q)]_{x \in R_0^+} = \begin{cases} q - p + r & \text{se } r > 0 \wedge q - p > 0 \\ q - p & \text{se } r \leq 0 \wedge q - p > 0 \\ 0 & \text{se } q - p \leq 0 \end{cases}$$

2.4.4 Considerazioni sui Content and Service Provider (CSP)

Il CSP, come già anticipato, ottiene ricavi da due diverse fonti:

1. Dal prezzo di utilizzo del servizio pagato dai consumatori: $P(x) = p * x$.
2. Dalla vendita dei dati degli utenti: $D(x) = d * x$.

È chiaro che il CSP tiene conto, nel fissare il prezzo, che un prezzo più basso ai consumatori, aumenta l'utilizzo del servizio, incrementando la quantità di dati da loro rilasciati che possono essere successivamente monetizzati. Non è raro, infatti, che molti servizi digitali abbiano un prezzo pari a zero o negativo (es. Facebook o Satispay).

Per poter ricavare dai dati il CSP potrebbe venderli per realizzare pubblicità personalizzate oppure per effettuare analisi su categorie di consumatori (ad esempio, esiste un mercato dei dati di semplice localizzazione).

Nel modello si ipotizza che il guadagno derivante dalla vendita di dati da parte dei CSP sia pari a $D(x)$ quando il consumatore utilizza il servizio a un determinato livello x .

Se nel mercato risulta presente anche il PDB, egli esercita una pressione competitiva al CSP e di conseguenza i loro ricavi sono ridotti di un fattore $\delta \in (0, 1]$.

- Se δ è basso, i ricavi dei CSP sono fortemente ridotti (δ prossimo a 0) di conseguenza la pressione competitiva esercitata dal PDB è molto alta. In questo caso i dati non hanno particolare valore esclusivo, poiché anche il PDB è in grado di sfruttarli più efficacemente.
- Se δ è alto, i ricavi dei CSP rimangono pressoché invariati (δ prossimo a 1) questo significa che la capacità del CSP di ottenere ricavi dai dati non è influenzata dalla capacità o meno del PDB, dunque i dati del CSP sono molto valorosi solo se usati dal CSP e non da altri attori.

Si può concludere che δ rappresenta l'efficienza del CSP nel monetizzare i dati se è presente il PDB. Senza la presenza del PDB si assume il CSP non subisce riduzione di ricavi da concorrenza ($\delta = 1$).

Nel modello si assume una relazione lineare tra utilizzo e ricavi dai dati:

$$D(x) = d * x$$

I profitti del CSP senza PDB sono pari a:

$$\pi_{CSP_{senza\ PDB}} = D(x) + P(x) - C(q)$$

Nota: $C(q)$ è una funzione di costo che rappresenta i costi sostenuti dal CSP per offrire il servizio ad un certo livello di qualità q . Si assume essere pari a $C(q) = \frac{q^2}{2}$

I profitti del CSP con PDB sono pari a:

$$\pi_{CSP_{con\ PDB}} = \delta * D(x) + P(x) - C(q)$$

2.4.5 Considerazioni sui Personal Data Broker (PDB)

Il PDB ha come unica fonte di ricavo ciò che riesce a ottenere dalla monetizzazione dei dati concessi dal consumatore. Il PDB, come già dichiarato, può incentivare i consumatori a trasferire i propri dati pagando una ricompensa $R(x) = r * x$ con $r > 0$. È importante, per studiare il modello, che sia il CSP sia il PDB dispongano degli stessi strumenti di prezzo per incentivare l'utilizzo del servizio e la generazione di dati.

Il PDB può ottenere un ricavo più alto o più basso rispetto al CSP in base alla sua capacità di monetizzare i dati degli utenti. Egli, infatti, potrebbe avere accesso a ulteriori set di dati, complementari a quelli trasferiti dal CSP, e quindi generare un maggior valore. Al contrario, se i dati trasferiti mancano di un contesto chiaro e definito, non sono così informativi e quindi difficili da monetizzare per il PDB.

Per consentire queste relazioni si introduce il parametro $\varrho \in [0, \infty)$ tale per cui se $\varrho = 1$ PDB e CSP sono ugualmente efficienti, se $\varrho < 1$ il CSP è più efficiente del PDB, per $\varrho > 1$ il PDB è più efficiente del CSP.

Si può affermare dunque che ϱ indica l'efficienza del PDB nella monetizzazione dei dati rispetto al CSP.

Il profitto del PDB è influenzato sia dalla qualità dei dati forniti dal CSP (si utilizzerà il termine riferito alla vendita dei dati nella funzione di profitto dei CSP in presenza del PDB: $\delta * D(x)$) e dalla sua capacità di estrarre informazioni; pertanto, la funzione di profitto è

$$\pi_{PDB} = \varrho * \delta * D(x) - R(x)$$

È importante ricordare che, qualora: $\delta * D(x) + \varrho * \delta * D(x) > D(x)$ l'introduzione del PDB nel mercato aumenta il valore totale generato dai dati creando un surplus economico. Questo accade se:

$$\delta * D(x) + \varrho * \delta * D(x) > D(x) \rightarrow \delta + \varrho * \delta > 1 \rightarrow \varrho > \frac{1 - \delta}{\delta}$$

2.4.6 Equilibrio senza PDB

Procedendo per backward induction, tralasciando lo stadio del gioco 3, poiché il PDB è assente, si procede dallo stadio 2 andando a determinare il prezzo p , una volta determinato p si passa allo stadio 1 del gioco andando a determinare la qualità del servizio q .

Si sa che in assenza di PDB $r = 0$ dunque $x(q, p) = q - p$.

La funzione di profitto del CSP è pari a

$$\pi CSP_{senza PDB} = D(x) + P(x) - C(q) \text{ con } C(q) = \frac{q^2}{2}$$

Può essere scritta come:

$$\pi CSP_{senza PDB} = d * x + p * x - \frac{q^2}{2} = d(q - p) + p(q - p) - \frac{q^2}{2}$$

Applicando la condizione di primo ordine rispetto al prezzo si ottiene:

$$\frac{d\pi CSP_{senza PDB}}{dp} = 0 \rightarrow -d + q - 2p = 0 \rightarrow p = \frac{q - d}{2}$$

Ottenuto il prezzo, lo si sostituisce all'interno della funzione di profitto e applicando la condizione di primo ordine rispetto alla qualità si ottiene:

$$\begin{aligned} \pi CSP_{senza PDB} &= dq - d \frac{q - d}{2} + \frac{q - d}{2} q - \left(\frac{q - d}{2} \right)^2 - \frac{q^2}{2} \\ \frac{\pi CSP_{senza PDB}}{dq} &= 0 \rightarrow d - \frac{d}{2} + q - \frac{d}{2} - q + d - q = 0 \rightarrow q = d \end{aligned}$$

In conclusione, sostituendo il valore di q all'interno dell'espressione del p si ricava:

$$p = 0$$

$$x = d$$

2.4.7 Commento al modello senza PDB

Vista l'assenza del PDB, il CSP vuole monetizzare il più possibile i dati degli utenti. Egli è quindi fortemente incentivato a fissare un prezzo di utilizzo basso per i consumatori al fine di stimolarne l'uso e di conseguenza la generazione di molti più dati.

Va specificato che il risultato $p = 0$ non è generale ma è frutto della funzione di domanda. Nonostante questo, è pienamente dimostrato dalla letteratura che i CSP sono chiaramente

incentivati a fissare un prezzo più basso per i consumatori in assenza di concorrenza da parte del PDB.

Esempi concreti possono essere Facebook, WhatsApp o Satispay. Nel caso dei primi due, il servizio è noto per essere gratuito. I loro ricavi derivano principalmente dalla pubblicità e dall'analisi dei dati degli utenti. Nel caso di Satispay, il prezzo può essere addirittura “negativo”: spesso, l'applicazione offre, al momento dell'iscrizione, accrediti di denaro agli utenti. Le fonti di guadagno per Satispay non derivano solo dal lato dei commercianti, ma anche dalla raccolta di informazioni sulle transazioni effettuate dagli utenti, tra cui: dettagli del pagamento, dispositivo utilizzato, geolocalizzazione e indirizzi IP.

2.4.8 Equilibrio con PDB

In presenza del PDB la funzione di domanda si è ricavato essere: $x(q, p, r) = q - p + r$.

Nello stadio 2 del gioco il CSP determina p mentre il PDB determina r .

Si hanno le seguenti funzioni di profitto:

$$\pi_{CSP_{con\ PDB}} = \delta * D(x) + P(x) - C(q) = \delta dx + px - \frac{q^2}{2}$$

$$\pi_{PDB} = q * \delta * D(x) - R(x) = q\delta dx - rx$$

Sostituendo in entrambe le funzioni di profitto x con $q - p + r$ si ottiene:

$$\pi_{CSP_{con\ PDB}} = \delta d(q - p + r) + p(q - p + r) - \frac{q^2}{2}$$

$$\pi_{PDB} = q\delta d(q - p + r) - r(q - p + r)$$

Applicando le condizioni di primo ordine alla prima rispetto a p e alla seconda rispetto a r si ottiene:

$$\frac{d\pi_{CSP_{con\ PDB}}}{dp} = 0 \rightarrow -\delta d + q - 2p + r = 0 \rightarrow p = \frac{1}{2}(q - \delta d + r)$$

$$\frac{d\pi_{PDB}}{dr} = 0 \rightarrow q\delta d - q + p - 2r = 0 \rightarrow r = \frac{1}{2}(q\delta d + p - q)$$

Nota: per quanto riguarda la ricompensa r , somma pagata da parte del PDB all'utente per monetizzare sui loro dati, deve essere positiva, e quindi $p \geq q - q\delta d$. Se così non fosse si verificherebbe una situazione paradossale: l'utente andrebbe a pagare il PDB che poi monetizzerebbe con i suoi dati. Di conseguenza se quella disequazione non è rispettata, si impone una ricompensa minima positiva infinitesimale r pari a ϵ con $\epsilon \rightarrow 0$. Questa scelta garantisce che gli utenti partecipino, ma senza generare un reddito effettivamente significativo per loro e per il PDB.

Dunque, si scrive:

$$r(p, q) = \begin{cases} \frac{1}{2}(\varrho\delta d + p - q) & \text{se } p \geq q - \varrho\delta d \\ \epsilon & \text{altrimenti} \end{cases}$$

$$p(r, q) = \frac{1}{2}(q - \delta d + r)$$

L'equilibrio è dato dall'incrocio delle due equazioni. Ovviamente, avendo due distinte definizioni di r , si otterranno, come precedentemente anticipato, 2 equilibri.

Utilizzando la prima equazione per il valore di $r(p, q)$ si ottiene:

$$\begin{cases} r = \frac{1}{2}(\varrho\delta d + p - q) \\ p = \frac{1}{2}(q - \delta d + r) \end{cases}$$

Sostituendo si ottiene

$$r = \frac{1}{2}\left(\varrho\delta d + \frac{1}{2}(q - \delta d + r) - q\right) \rightarrow r = \frac{1}{2}\varrho\delta d + \frac{1}{4}q - \frac{1}{4}\delta d + \frac{1}{4}r - \frac{1}{2}q \rightarrow$$

$$\frac{3}{4}r = \frac{1}{2}\varrho\delta d - \frac{1}{4}\delta d - \frac{1}{4}q \rightarrow r = \frac{1}{3}(\delta d(2\varrho - 1) - q)$$

Effettuando le apposite sostituzioni si ricava:

$$p = \frac{1}{3}(\delta d(\varrho - 2) + q)$$

L'equilibrio composto da $p = \frac{1}{3}(\delta d(\varrho - 2) + q)$ e da $r = \frac{1}{3}(\delta d(2\varrho - 1) - q)$ è l'equilibrio di reddito positivo.

Utilizzando la seconda equazione per il valore di $r(p, q)$ si ottiene:

$$\begin{cases} r = \epsilon \\ p = \frac{1}{2}(q - \delta d + r) \end{cases} \rightarrow \begin{cases} r = \epsilon \\ p = \frac{1}{2}(q - \delta d + \epsilon) \end{cases}$$

Questo è l'equilibrio di reddito minimo.

Riassumendo:

Equilibrio di reddito positivo	Equilibrio di reddito minimo
$p = \frac{1}{3}(\delta d(\varrho - 2) + q)$	$p = \frac{1}{2}(q - \delta d + \epsilon)$
$r = \frac{1}{3}(\delta d(2\varrho - 1) - q)$	$r = \epsilon$

Allo stadio 1 del gioco, il CSP effettua l'investimento per determinare la qualità del servizio che vuole erogare massimizzando il proprio profitto.

$$\pi_{CSP_{con PDB}} = \delta d(q - p + r) + p(q - p + r) - \frac{q^2}{2}$$

Equilibrio di reddito positivo:

$$p = \frac{1}{3}(\delta d(\varrho - 2) + q) \text{ e } r = \frac{1}{3}(\delta d(2\varrho - 1) - q)$$

Andando a sostituire nella formula del profitto il valore di p e r si ottiene:

$$\begin{aligned} \pi = \delta d \left[q - \frac{1}{3}(\delta d(\varrho - 2) + q) + \frac{1}{3}(\delta d(2\varrho - 1) - q) \right] \\ + \left[\frac{1}{3}(\delta d(\varrho - 2) + q) \right] \left[q - \frac{1}{3}(\delta d(\varrho - 2) + q) + \frac{1}{3}(\delta d(2\varrho - 1) - q) \right] \\ - \frac{q^2}{2} \end{aligned}$$

Effettuando i calcoli, si ottiene:

$$\pi_{CSP_{con PDB}}(p_+, r_+, q) = \frac{1}{18}(2d^2\delta^2(\varrho + 1)^2 + 4d\delta(\varrho + 1)q - 7q^2)$$

Equilibrio di reddito minimo:

$$p = \frac{1}{2}(q - \delta d + \epsilon); r = \epsilon$$

$$\pi = \delta d \left[q - \frac{1}{2}(q - \delta d + \epsilon) + \epsilon \right] + \left[\frac{1}{2}(q - \delta d + \epsilon) \right] \left[q - \frac{1}{2}(q - \delta d + \epsilon) + \epsilon \right] - \frac{q^2}{2}$$

Effettuando i calcoli si ottiene:

$$\pi_{CSP_{con PDB}}(p_{min}, r_{min}, q) = \frac{1}{4}[(d\delta + \epsilon)^2 + 2q(d\delta + \epsilon) - q^2]$$

Riassumendo:

Equilibrio di reddito positivo	Equilibrio di reddito minimo
$p = \frac{1}{3}(\delta d(\varrho - 2) + q)$	$p = \frac{1}{2}(q - \delta d + \epsilon)$
$r = \frac{1}{3}(\delta d(2\varrho - 1) - q)$	$r = \epsilon$
$\pi = \frac{1}{18}(2d^2\delta^2(\varrho + 1)^2 + 4d\delta(\varrho + 1)q - 7q^2)$	$\pi = \frac{1}{4}[(d\delta + \epsilon)^2 + 2q(d\delta + \epsilon) - q^2]$

Giunti a questo punto si può determinare la qualità q ottimale nei due scenari.

Nel caso di equilibrio di reddito positivo:

$$\frac{d\pi_{CSP \text{ con } PDB}(p_+, r_+, q)}{dq} = 0 \rightarrow \frac{4}{18}d\delta(\varrho + 1) - \frac{14}{18}q = 0 \rightarrow q_+ = \frac{2}{7}d\delta(\varrho + 1)$$

Nel caso di equilibrio di reddito minimo:

$$\frac{d\pi_{CSP \text{ con } PDB}(p_{min}, r_{min}, q)}{dq} = 0 \rightarrow \frac{1}{2}(d\delta + \epsilon) - \frac{1}{2}q = 0 \rightarrow q_{min} = d\delta + \epsilon$$

Riassumendo:

Equilibrio di reddito positivo	Equilibrio di reddito minimo
$p = \frac{1}{3}(\delta d(\varrho - 2) + q)$	$p = \frac{1}{2}(q - \delta d + \epsilon)$
$r = \frac{1}{3}(\delta d(2\varrho - 1) - q)$	$r = \epsilon$
$\pi = \frac{1}{18}(2d^2\delta^2(\varrho + 1)^2 + 4d\delta(\varrho + 1)q - 7q^2)$	$\pi = \frac{1}{4}[(d\delta + \epsilon)^2 + 2q(d\delta + \epsilon) - q^2]$
$q = \frac{2}{7}d\delta(\varrho + 1)$	$q = d\delta + \epsilon$

Una volta determinata la q ora si può procedere nel determinare tutti i vari parametri dei modelli.

Nel caso di equilibrio di reddito positivo:

$$q_+ = \frac{2}{7}d\delta(\varrho + 1)$$

$$p_+ = \frac{1}{3}(\delta d(\varrho - 2) + q) = \frac{1}{3}\left(\delta d(\varrho - 2) + \frac{2}{7}d\delta(\varrho + 1)\right) = \frac{1}{3}\left(\delta d\varrho - 2\delta d + \frac{2}{7}d\delta\varrho + \frac{2}{7}d\delta\right) = \frac{1}{3}\left(\frac{9}{7}\delta d\varrho - \frac{12}{7}d\delta\right) = \frac{1}{7}(3\delta d\varrho - 4d\delta) = \frac{1}{7}d\delta(3\varrho - 4)$$

$$p_+ = \frac{1}{7}d\delta(3\varrho - 4)$$

$$r_+ = \frac{1}{3}(\delta d(2\varrho - 1) - q) = \frac{1}{7}d\delta(4\varrho - 3)$$

$$x_+ = q - p + r = \frac{3}{7}d\delta(\varrho + 1)$$

$$\pi_{CSP+} = \frac{1}{18}(2d^2\delta^2(\varrho + 1)^2 + 4d\delta(\varrho + 1)q - 7q^2) = \frac{1}{7}d^2\delta^2(\varrho + 1)^2$$

$$\pi_{PDB} = \varrho\delta d(q - p + r) - r(q - p + r) = \frac{9}{49}(\delta d + \delta d\varrho)^2$$

Nel caso di equilibrio di reddito minimo:

$$q_{min} = d\delta + \epsilon$$

$$p_{min} = \frac{1}{2}(q - \delta d + \epsilon) = \epsilon$$

$$r_{min} = \epsilon$$

$$x_{min} = q - p + r = d\delta + \epsilon - \epsilon + \epsilon = d\delta + \epsilon$$

$$\pi_{CSPmin} = \frac{1}{4}[(d\delta + \epsilon)^2 + 2q(d\delta + \epsilon) - q^2] = \frac{1}{2}(d\delta + \epsilon)^2$$

$$\pi_{PDB min} = \varrho\delta d(q - p + r) - r(q - p + r) = (d\delta + \epsilon)(d\delta\varrho - \epsilon)$$

Infine, si procede nell'andare a definire il valore di ϱ , capacità di monetizzare i dati da parte del PDB, nelle due situazioni di equilibrio.

È chiaro che, nella condizione di reddito positivo, deve verificarsi che $r_+(q_+) \geq \epsilon$ e che quindi:

$$\frac{1}{7}d\delta(4\varrho - 3) > \epsilon \rightarrow \varrho > \frac{3d\delta + 7\epsilon}{4d\delta} = \bar{\varrho}_+$$

Va da sé che il valore di ϱ deve essere maggiore o uguale a $\bar{\varrho}_+$ per la realizzazione dell'equilibrio di reddito positivo: $\varrho \geq \bar{\varrho}_+$

L'equilibrio di reddito minimo si verifica se $r_+(q_{min}) < \epsilon$ e dunque:

$$\frac{1}{3}(\delta d(2q - 1) - d\delta - \epsilon) < \epsilon \rightarrow \delta d 2q - \delta d - d\delta - \epsilon < 3\epsilon \rightarrow q < \frac{d\delta + 2\epsilon}{d\delta} = \bar{q}_{min}$$

Va da sé che il valore di q deve essere minore di $\bar{q}_{min}$ per la realizzazione dell'equilibrio minimo: $q < \bar{q}_{min}$

Infine, è chiaro che il CSP preferisce l'equilibrio di reddito positivo se il π_{CSP+} è maggiore rispetto a π_{CSPmin} ovvero:

$$\frac{1}{7}d^2\delta^2(q + 1)^2 > \frac{1}{2}(d\delta + \epsilon)^2$$

Risolvendo questa disequazione per il valore di efficienza del PDB q si ottiene:

$$(q + 1)^2 > \frac{7}{2} * \frac{(d\delta + \epsilon)^2}{d^2\delta^2} \rightarrow q > \sqrt{\frac{7}{2} * \frac{d^2\delta^2 + 2d\delta + \epsilon^2}{d^2\delta^2}} - 1 = \bar{q}$$

Si indica questo valore con $\bar{q}$. Per tutti i valori di $q > \bar{q}$ l'equilibrio di reddito positivo è più profittevole per il CSP.

Si può scrivere in maniera più formale la qualità dell'investimento che può effettuare il CSP allo stadio 1:

$$q = \begin{cases} q_+ = \frac{2}{7}d\delta(q + 1) & \text{se } q > \bar{q} \\ q_{min} = d\delta + \epsilon & \text{se } q < \bar{q} \end{cases}$$

2.4.9 Riassunto risultati matematici del modello con PDB

Equilibrio di reddito positivo $\varrho > \bar{\varrho}$	Equilibrio di reddito minimo $\varrho \leq \bar{\varrho}$
$\bar{\varrho} = \sqrt{\frac{7}{2}} * \sqrt{\frac{d^2\delta^2 + 2d\delta + \epsilon^2}{d^2\delta^2}} - 1$	
$q_+ = \frac{2}{7}d\delta(\varrho + 1)$	$q_{min} = d\delta + \epsilon$
$p_+ = \frac{1}{7}d\delta(3\varrho - 4)$	$p_{min} = \epsilon$
$r_+ = \frac{1}{7}d\delta(4\varrho - 3)$	$r_{min} = \epsilon$
$x_+ = \frac{3}{7}d\delta(\varrho + 1)$	$x_{min} = d\delta + \epsilon$
$\pi_{CSP+} = \frac{1}{7}d^2\delta^2(\varrho + 1)^2$	$\pi_{CSPmin} = \frac{1}{2}(d\delta + \epsilon)^2$
$\pi_{PDB+} = \frac{9}{49}(\delta d + \delta d\varrho)^2$	$\pi_{PDBmin} = (d\delta + \epsilon)(d\delta\varrho - \epsilon)$
$\bar{\varrho}_+ = \frac{3d\delta + 7\epsilon}{4d\delta}$	$\bar{\varrho}_{min} = \frac{d\delta + 2\epsilon}{d\delta}$

Tabella 1: I risultati dei modelli

2.4.10 Commento del modello con PDB

Innanzitutto, si intende ribadire che all'interno di un mercato con un CSP monopolistico e un PDB, si può verificare uno dei due tipi di equilibrio a seconda dell'efficienza del PDB nella generazione di ricavi dai dati.

Nell'equilibrio di reddito minimo, la somma offerta dal PDB al consumatore è appena sufficiente da far cedere i propri dati ma garantisce loro un reddito praticamente nullo dalla loro vendita.

A livello pratico questo avviene spesso sotto forma di coupon o buoni sconto personalizzati. Basta pensare, ad esempio, a quante volte applicazioni come Trenitalia o Italo inviano, tramite email, sondaggi o domande da compilare. Dopo aver fornito le risposte adeguate, il consumatore solitamente riceve sconti o agevolazioni sui prossimi acquisti. Questo corrisponde al concetto di ricompensa minima per l'utente, il quale fornisce dati senza ricevere però un reddito monetario significativo.

Bisogna però, porre attenzione ad un particolare: il prezzo netto che il consumatore valuta per l'utilizzo del servizio è pari a p (prezzo deciso dal CSP) $- r$ (ricompensa data dal PDB), in caso di equilibrio minimo $r \cong 0$ e di conseguenza è uguale a quello proposto senza il PDB. La differenza però è che il CSP, subendo la concorrenza del PDB offrirà un servizio di qualità peggiore rispetto ad una situazione senza PDB.

Nell'equilibrio di reddito positivo invece, la ricompensa fornita dal PDB offre ai consumatori un reddito strettamente positivo: $\frac{1}{7}d\delta(4q - 3)$. Più è alta l'efficienza q del PDB, maggiore sarà la ricompensa pagata ai consumatori. È ovvio che all'aumentare della ricompensa il prezzo netto pagato dagli utenti per il servizio diminuisce. Come si evince dalle formule la qualità del servizio dipende dall'efficienza del PDB, $\frac{2}{7}d\delta(q + 1)$, tuttavia, può comunque essere inferiore a quella senza PDB ($q = d$).

Di conseguenza i consumatori, nell'equilibrio di reddito positivo, traggono beneficio solo se il PDB risulta molto efficiente nel generare i ricavi dai dati, una situazione che nella pratica risulta particolarmente improbabile in quanto strutturalmente insostenibile.

Riprendendo l'esempio di Satispay in cui l'azienda svolge il compito sia di CSP (fornitrice del servizio) che PDB (raccolge e mette a disposizione i propri dati per eventuali commercializzazioni) la ricompensa esiste ma è quasi simbolica. L'utente riceve un beneficio economico, tra l'altro lo riceve una volta sola nella vita dell'applicazione,

che è irrilevante rispetto al valore che Satispay può ricavare dall'ammontare delle transazioni.

Un altro esempio che si può citare è quello tra Google (CSP) e Google Opinion Rewards (PDB). Quest'ultima è un'applicazione di sondaggi e inchieste, supportata da sistema Android e sviluppata dal team di Google Surveys.

L'app permette di ottenere dei crediti da utilizzare su Google Play (spendibili ad esempio per app, musica, giochi...) in cambio della compilazione da parte del consumatore di domande o questionari spesso da condividere con altre persone.

Il consumatore riceve una cifra compresa tra 0,10 \$ e 1 \$¹³ per sondaggio completato ma è importante specificare che non è possibile compilare infiniti sondaggi liberamente. Infatti, si può rispondere ad un sondaggio solamente dopo aver ricevuto la relativa notifica da Google.

Google sostiene che per ricevere più sondaggi convenga rispondere in modo “onesto e con informazioni corrette”, rimane oscura la modalità con cui Google valuta la correttezza o meno delle informazioni fornite.

È importante sottolineare come la ricompensa fornita da Google Opinion Rewards sia minima se confrontata con il valore che Google può estrarre dai dati raccolti. Ancora più rilevante è il fatto che i crediti ricevuti dagli utenti rimangono vincolati all'ecosistema Google: i soldi che sono accreditati vanno spesi per acquisti su Google Play e non possono essere spesi liberamente in altre piattaforme.

Non solo Google fa ritornare il denaro al suo interno ma per di più, attraverso gli acquisti che l'utente farà con quei crediti, ottiene ulteriori informazioni sui suoi gusti e comportamenti che possono tradursi in:

- Sondaggi sempre più mirati.
- Possibilità di applicare discriminazioni di prezzo: ad esempio, se Google osserva che un consumatore spende regolarmente per scaricare una determinata tipologia di libri, con il passare del tempo, potrebbe aumentare il credito richiesto per libri

¹³ Guida di Opinion Rewards

<https://support.google.com/opinionrewards/answer/7378183?hl=it#zippy=%2Cquanto-viene-pagata-la-partecipazione-a-ogni-sondaggio%2Cperch%C3%A9-non-ho-ricevuto-il-pagamento-per-un-sondaggio%2Cesiste-un-modo-per-rispondere-a-un-maggior-numero-di-sondaggi-per-guadagnare-di-pi%C3%B9%2Ccome-faccio-a-trasferire-i-miei-pagamenti-da-google-opinion-rewards-a-paypal>

dello stesso genere, costringendo il consumatore a compilare un numero sempre più crescente di sondaggi, che rilasciano tantissime informazioni monetizzabili da Google.

Si può concludere che la ricompensa erogata agli utenti è simbolica e vincolata all'ecosistema Google mentre il valore estratto dai dati è enormemente superiore. Il CSP e il PDB massimizzano i loro profitti e riducono la quota trasferita ai consumatori. L'utente finale accetta la condizione e percepisce la ricompensa come un vantaggio gratuito ma in realtà rimane strutturalmente in una posizione svantaggiosa. È ovvio che si finisce in una posizione di equilibrio di reddito minimo.

Un'altra osservazione che è bene ribadire riguarda un concetto che è stato descritto più volte durante la spiegazione del modello: in presenza del PDB la qualità del servizio offerta dal CSP scende.

Infatti, guardando la qualità in caso di equilibrio a reddito minimo è pari a $q_{min} = d\delta + \epsilon$ con $d \in (0,1]$ e $\epsilon \rightarrow 0$ è evidente essere minore rispetto a quella del modello senza PDB (d).

Guardando invece la qualità in caso di equilibrio a reddito positivo osserviamo immediatamente che:

$$\frac{dr}{dq} = \frac{d \frac{1}{3} (\delta d(2q - 1) - q)}{dq} < 0$$

Dunque, all'aumentare della ricompensa fornita dal PDB per il consumatore diminuisce la qualità del servizio fornita dal CSP.

Infine, nell'equilibrio di reddito positivo, c'è solo un caso in cui il CSP potrebbe essere incentivato a migliorare la qualità del servizio facendo leva sulla variabile prezzo. Questo accade perché, grazie alla ricompensa fornita dal PDB ai consumatori, il CSP può aumentare il prezzo del servizio senza ridurre l'uso da parte degli utenti. Poiché i dati forniti dagli utenti devono essere di buona qualità per ottenere ricompense maggiori dal PDB, il CSP può reinvestire i ricavi aggiuntivi derivanti dall'aumento del prezzo in ulteriori miglioramenti della qualità del servizio.

Teoricamente questo effetto funziona, il problema è che molti servizi che abbiamo citato sono gratuiti (Facebook, WhatsApp, Amazon, Satispay) e molto probabilmente se si

mettesse un prezzo o si aumenterebbe, molti utenti non sarebbero disposti a pagare, di conseguenza nella pratica sembra un effetto irrealizzabile.

Con gli effetti precedentemente descritti possiamo analizzare brevemente l'andamento dei profitti del PDB e del CSP.

Per quanto concerne il profitto del PDB è crescente con la sua efficienza nel saper ricavare informazioni dai dati q . Si ha una leggera diminuzione nel passaggio tra equilibrio di reddito minimo e quello di reddito positivo perché nella prima parte, la ricompensa fornita ai consumatori è pari a $\varepsilon \rightarrow 0$ e ricava dalla vendita dei dati. Con il passaggio al secondo equilibrio, il PDB deve aumentare la ricompensa e di conseguenza porterà ad una iniziale erosione dei profitti.

Per quanto riguarda i profitti del CSP nell'equilibrio di reddito minimo, soffrono dell'effetto concorrenza, ma dipendono da δ e non da q . Nell'equilibrio di reddito positivo invece, iniziano a dipendere da q e possono crescere a causa dell'effetto di appropriazione del valore dei dati.

Infine, si vuole effettuare una breve valutazione riguardante il surplus dei consumatori. Questo è misurato basandosi sulla domanda dei consumatori in quanto un utilizzo maggiore del servizio da parte degli utenti implica un surplus più alto da parte loro. Dal modello si evince che, come già ampiamente discusso, i consumatori nell'equilibrio di reddito minimo stanno sempre peggio rispetto ad uno scenario senza PDB, infatti:

$$x_{senza\ PDB} = d > x_{min} = d\delta + \epsilon$$

Nell'equilibrio di reddito positivo, invece si ha che:

$$x_+ = \frac{3}{7}d\delta(q + 1)$$

E dunque possiamo ricavare che:

$$x_+ > x_{senza\ PDB} \rightarrow \frac{3}{7}d\delta(q + 1) > d \rightarrow q > \frac{7 - 3\delta}{3\delta}$$

Dunque, il surplus è maggiore solamente se $q > \frac{7-3\delta}{3\delta}$ il che implica un PDB particolarmente efficiente ma per tutte le ragioni che abbiamo sopra descritto è una situazione oggettivamente complicata da raggiungere.

Al termine dell'analisi, si può affermare che l'idea dei PDB come redistributori del valore nel mercato dei dati non risulta del tutto soddisfacente. Nella pratica, infatti le ricompense

sono generalmente troppo basse per generare un reale beneficio, e il mercato rimane dominato dalle grandi aziende che generano e gestiscono i dati.

2.5 I principali attori nel mercato

In questo paragrafo sono riportati i nomi, con una breve descrizione, delle più importanti aziende che hanno a che fare con il mercato dei dati.

Andando a riprendere la distinzione effettuata nel paragrafo precedente che distingue i CSP dai PDB l'articolo realizzato da Wolfie Christl, *Corporate Surveillance in Everyday Life* riporta la seguente classificazione.

Tra i CSP si trovano le piattaforme online tra cui:

- Meta che comprende Facebook, WhatsApp e Instagram. Tutti e tre gestiscono miliardi di utenti
- Google che con tutti i sistemi Android, utilizzatori di Gmail, visualizzatori di YouTube e molti altri servizi possiede dati su miliardi di persone
- Apple che utilizza tutti i dati dei dispositivi iOS per raccogliere informazioni in maniera continuativa
- Microsoft: essendo diffuso su gran parte dei computer a livello globale, raccoglie costantemente dati e si colloca al diciassettesimo posto tra le più grandi media company
- Amazon Web Services: permette di ricavare numerosissime informazioni sulle preferenze e sui gusti dei consumatori
- Walt Disney: raccoglie, grazie alla propria piattaforma, dati relativi alle preferenze cinematografiche degli utenti.
- Bertelsmann: azienda leader nel settore dei media, con sede in Germania, dei servizi e della formazione nel settore dei dati e nell'intelligenza artificiale.
- Viacom: gruppo multimediale statunitense del settore della comunicazione dell'intrattenimento
- CBS: importante radiotelevisiva statunitense
- Gruppo Globo: conglomerato di società brasiliane concentrate in particolar modo nel settore dei media e della comunicazione
- Spotify: raccoglie dati sull'utilizzo dell'applicazione da parte degli utenti, captando informazioni come il loro mood, le loro attività e la geolocalizzazione

Un importante ruolo nella raccolta dati lo svolgono le società finanziarie, agenzie di credito e di compagnie assicurative tra cui:

- PayPal: azienda più importante per quanto concerne i pagamenti online e condivide informazioni personali con più di 600 terze parti, tra cui: altre società di pagamento, agenzie di credito, aziende di verifica dell'identità e rilevamento frodi.
- Experian, Equifax e TransUnion: sono le tre principali agenzie di credito a livello mondiale. Raccolgono, memorizzano e analizzano informazioni personali degli individui, tra cui: prestiti, operazioni con carte di credito, attitudini al pagamento ecc. ecc. Essi hanno dati su circa un miliardo di persone.
- MasterCard e Visa: la prima associa i propri dati con popolazioni di cookie, in modo da collegare le spese alle attività online di un individuo. Gestisce circa 95 miliardi di transazioni reali provenienti da 2 miliardi di titolari di carte in 210 paesi in tutto il mondo. La seconda invece combina i propri dati transazionali con dati demografici e altri dati provenienti da Oracle.
- Progressive Corporation: è una delle principali compagnie assicurative statunitensi, che oltre ad applicare polizze tradizionali, riesce a raccogliere e analizzare dati sui clienti in maniera nettamente superiore rispetto alle altre aziende. Questo è possibile sia grazie a numerosi dispositivi tecnologici installati nei veicoli che permettono di monitorare lo stile e le abitudini di guida e sia grazie ad una flotta di personale che, con i noti furgoncini, si reca immediatamente sul posto dell'incidente così da poter raccogliere ulteriori dati.

Altrettanto importanti risultano le aziende di telecomunicazioni e di fornitura servizio internet tra le altre: AT&T, Comcast, Verizon, Vodafone, Telecom. Questi sono stati colti più volte ad inserire meccanismi di tracciamento nelle attività di navigazione web dei propri utenti: i super cookie, che permettono la raccolta di dati sugli utenti poi utilizzati a scopo pubblicitario e di monitoraggio.

Per quanto concerne i principali Consumer Data Brokers troviamo:

- Acxiom detiene dati su circa 700 milioni di persone raccolti tramite miliardi di cookies ed è in grado di gestire circa 44 miliardi di profili.

- Circana: azienda americana che nasce dalla fusione tra IRI e NPD Group. Essa accede ai dati di oltre 85.000 negozi tra cui alimentari club, minimarket, negozi di liquori, di animali domestici.
- Merkle: afferma di gestire più di 3,7 miliardi di record di clienti dei propri clienti: es. Dell, Nespresso, American Express, Marriott.
- Arvato: mantiene relazioni con oltre 600 milioni di consumatori.

Si vuole ricordare infine che, oltre a CSP e PDB sono coinvolte anche le istituzioni politiche in ambiti come sanità, istruzione, sicurezza nazionale e per scopi di campagna elettorale.

Come schema riassuntivo si riporta la seguente immagine che presenta una mappatura del panorama commerciale del tracciamento e della profilazione digitale:

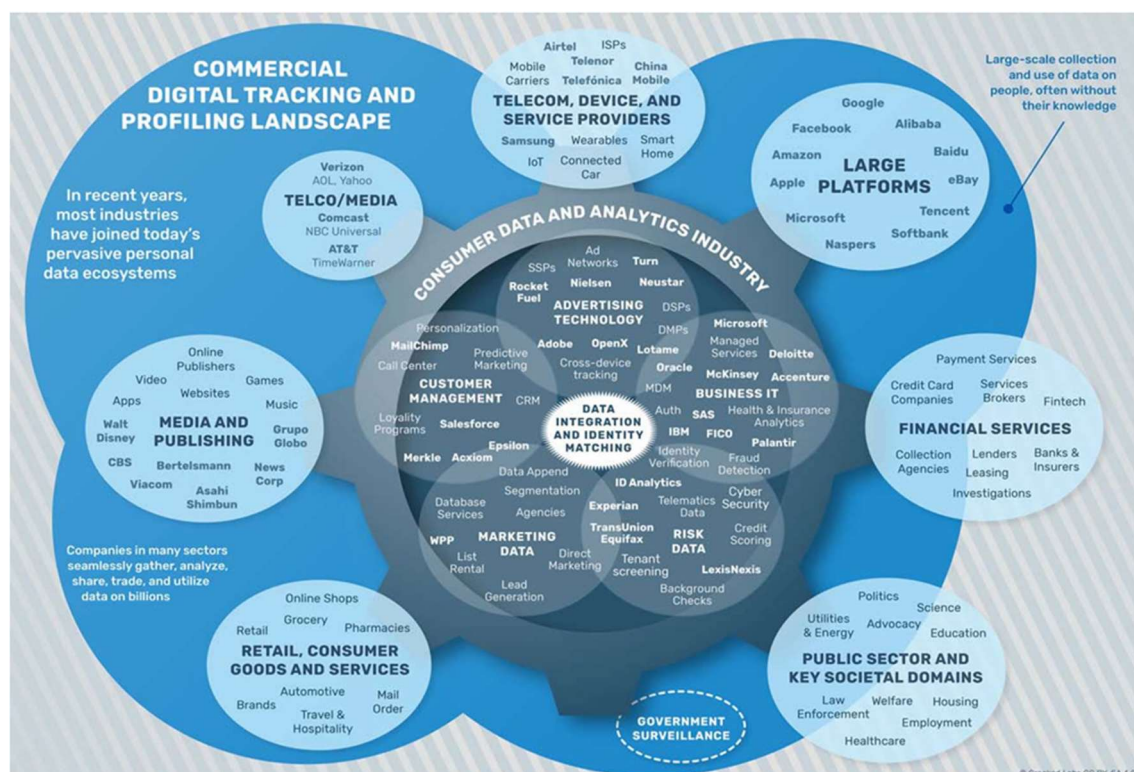


Figura 11: I principali attori - Corporate Surveillance in Everyday Life

2.6 Una tendenza del mercato: le acquisizioni

All'interno di questo paragrafo si vuole evidenziare come, sebbene l'introduzione dei PDB fosse volta a redistribuire il valore dei dati verso i consumatori in un mercato che rischiava di essere dominato dai CSP, si è verificato un effetto opposto. I PDB, infatti,

faticano a trasferire valore ai consumatori, come testimoniato dal modello; per di più, parallelamente, aumentano le acquisizioni tra società che gestiscono dati, con conseguente concentrazione di informazioni nelle mani di un numero sempre più ridotto di CSP. Queste operazioni spesso sono guidate da obiettivi strategici chiari, tra cui: ampliare e migliorare le capacità di analisi dei dati, rafforzare il vantaggio competitivo con la consolidazione di una potenza economica imbattibile e raggiungere un potere economico difficilmente gestibile dalle istituzioni.

Un primo esempio riguarda il grande rivenditore britannico Tesco. Si tratta di una grande catena inglese di supermercati, fondata nel 1919 in East London e poi cresciuta in 14 diversi paesi nel mondo, che ha trasferito le proprie attività legate alla carta fedeltà e alla gestione dei dati a Dunhumby (società da Tesco controllata) il cui slogan era *“trasformare i dati dei clienti in loro soddisfazione”*. Nel 2014 Dunhumby ha acquisito la società tedesca di pubblicità online Sociomantic, con l’obiettivo di combinare i propri dati sulle preferenze di acquisto di più di 400 milioni di consumatori con i dati in tempo reale di oltre 700 milioni di utenti online. Non è chiaro quali dati siano utilizzati, ma è ragionevole supporre che una parte significativa di queste informazioni provenga dai dati della carta fedeltà.

Non stupisce, vista la presenza sempre più massiccia degli smartphone nelle nostre vite, che anche le compagnie telefoniche sono protagoniste di questo processo.

Comcast, società statunitense nel campo delle telecomunicazioni (servizi telefonici, internet e tv) ha acquistato NBC Universal (operatore nel settore dei media) per 13,75 miliardi di dollari¹⁴, diventando un nuovo colosso del mondo dell’entertainment ai livelli di Disney. Questo ha permesso la creazione di un più vasto portafoglio di contenuti e, soprattutto, ha ampliato la pubblicità cross-device e le capacità di analisi dati.

Un caso simile riguarda l’acquisizione da parte di Verizon (fornitore statunitense di banda larga e di telecomunicazioni, ed è la più grande compagnia telefonica seguita da AT&T) di AOL e Yahoo per un importo superiore a 9 miliardi di dollari, con lo scopo di legare le informazioni ottenute dai motori di ricerca a quelli delle piattaforme digitali. Tuttavia, è

¹⁴ APCOM – New York – 03/12/2009

corretto sottolineare che entrambe nel 2021, a seguito della crisi pandemica, sono state cedute al fondo Apollo Global Management per 5 miliardi di dollari.

Un ulteriore esempio rilevante riguarda Telenor (compagnia norvegese di telecomunicazioni) che ha acquisito Tapad per 360 milioni di dollari, una delle principali società di tracciamento cross-device. Essa possiede dati su 2 miliardi di dispositivi in tutto il mondo. È lo stesso CEO e co-fondatore di Tapad ad affermare che dispongono di dati anonimizzati ricavati da dispositivi personali, oltre a numerose informazioni sui comportamenti d'acquisto e su ciò che i consumatori stanno facendo in tempo reale.

Nei tempi più recenti si registra l'acquisto di WhatsApp da parte di Facebook per circa 19 miliardi di euro nel 2014, quando l'app contava 450 milioni di utenti, oggi ha raggiunto quota un miliardo.

L'operazione ha suscitato un ampio dibattito pubblico, in particolare il Sole 24 Ore di giovedì 10/09/2015 titolava: *“Facebook si mangia i messaggini di WhatsApp: 19 miliardi \$ nel piatto di Zuckerberg”*¹⁵ inoltre secondo due organizzazioni no-profit americane, come l'Electronic Privacy Information Center e il Center for Digital Democracy, Zuckerberg ha comprato non gli utenti ma i loro numeri telefonici con le informazioni sensibili a loro annesse. Nonostante le polemiche sia la Federal Trade Commission statunitense e sia l'Unione Europea hanno approvato l'operazione. È evidente che una cifra così elevata sia probabilmente giustificata dalla prospettiva di trasformare WhatsApp in una fonte strategica di dati a supporto del core business pubblicitario di Facebook, anche se ufficialmente le due aziende hanno dichiarato di mantenere separati i modelli di business. Resta comunque da comprendere, però, una spesa così elevata per mantenere le aziende così com'erano.

Un'altra operazione di portata ancora maggiore è stato l'acquisto di LinkedIn da parte di Microsoft per circa 26 miliardi di dollari. Qui, il Wall Street Journal dichiara che LinkedIn manterrà il proprio brand distintivo, cultura e indipendenza.

¹⁵ Facebook si mangia i messaggini di WhatsApp: 19 miliardi di \$ nel piatto di Zuckerberg – Angelo Paura – 20/02/2014 – Il Sole 24 Ore - [https://st.ilsole24ore.com/art/tecnologie/2014-02-20/facebook-si-mangia-messaggini-WhatsApp-19-miliardi-\\$-piatto-zurckerberg-074244.shtml](https://st.ilsole24ore.com/art/tecnologie/2014-02-20/facebook-si-mangia-messaggini-WhatsApp-19-miliardi-$-piatto-zurckerberg-074244.shtml)

Tuttavia, si ritiene che l'integrazione di LinkedIn, in cui compaiono professionisti da tutto il mondo, con un sistema dominato da Microsoft 365, suggerisca possibili sinergie nella raccolta e nell'utilizzo dei dati che potrebbero, un giorno, influenzare lo sviluppo di prodotti e servizi con strategie di prezzo differenziate. Qualora così non fosse, anche in questo caso, risulterebbe complicato comprendere una spesa così onerosa.

L'ultimo esempio che si riporta riguarda l'annuncio dell'acquisizione da parte di Google della società israeliana di cloud e cybersecurity Wiz per 32 miliardi di dollari. Si tratta di una startup fondata nel 2020 da un team di ex militari dell'Unità 8200, divisione di intelligence tecnologica delle Forze di Difesa Israeliana con un fatturato annuo di 700 milioni di dollari. Si vuole far notare che, solitamente, per aziende normali il valore di vendita è stabilito utilizzando un fattore per 10, risulta quindi essere venduta a quasi 5 volte di più rispetto a quanto un'azienda normale sarebbe stata venduta.

L'operazione che consente a Google di entrare in competizione in settori in cui Amazon Web Services risulta più potente solleva questioni etiche significative: le tecnologie di Wiz sono state impiegate in attività di sorveglianza avanzata come il riconoscimento biometrico facciale ai checkpoint e il monitoraggio di attivisti e giornalisti. L'ingresso di tale strumento in un'azienda così popolare come Google solleva gravi preoccupazioni riguardo a privacy, diritti umani e libertà individuali.

In conclusione, questo paragrafo mette in evidenza come il mercato dei Big Data sia sempre più dominato da processi di concentrazione tramite acquisizioni finanziarie. Queste operazioni accrescono le capacità tecnologiche e competitive delle aziende, concentrano i dati e le informazioni, spesso sensibili, in mano a pochi attori e infine, questa dinamica rischia di ridurre la trasparenza e di compromettere il ruolo delle istituzioni nel garantire la tutela dei diritti fondamentali dei cittadini.

2.7 I dati raccolti:

Negli ultimi anni moltissime aziende hanno iniziato a monitorare, tenere traccia e registrare le persone, soprattutto per mezzo delle tecnologie, in ogni aspetto della loro vita. Ad essere analizzati sono: i comportamenti, gli spostamenti, le relazioni sociali, gli interessi, le debolezze e moltissimi altri elementi privati che sono costantemente registrati, valutati e analizzati in tempo reale. È importante chiarire sin da subito che a

questo monitoraggio non partecipano solamente le grandi applicazioni come Facebook o Google, ma anche moltissime altre aziende che, in base ai click degli utenti, registrano e si scambiano informazioni.

Gli smartphone sono oggi i principali strumenti di raccolta dati, essi forniscono approfondimenti dettagliati sulla personalità e sulla vita quotidiana di un utente. Inoltre, per funzionare al meglio, questi dispositivi necessitano di applicazioni come Google o Microsoft che, solamente attraverso il loro download intrecciano numerosissime informazioni personali.

La raccolta dati, tuttavia, non si limita ai telefoni cellulari e alle loro applicazioni ma coinvolge numerosi altri dispositivi che, tramite la connessione di rete, collezionano tantissime informazioni preziose, ad esempio: smart TV, contatori, termostati, antifurti, stampanti, frigoriferi, automobili, e-reader, assistenti vocali e tanto altro.

Numerosi studi scientifici dimostrano che molti aspetti della personalità di un utente possono essere dedotti dalle analisi sui dati relativi a: ricerche web, attività social, video visualizzati, acquisti o cronologie di navigazione. Da queste informazioni si possono ricavare dati estremamente sensibili: etnia, religione, orientamento politico e sessuale, consumo di alcol o droghe, condizioni psicologiche o grado di soddisfazione della vita.

Nell'immagine sottostante è riportata l'accuratezza con cui certe informazioni sono dedotte dai Like di Facebook.

Si nota come la percentuale di accuratezza riguardante l'orientamento politico è pari all'85%: è importante ricordare che una volta avvenuto il monitoraggio, esercitare una possibile influenza non è un pensiero così proibitivo...

È interessante notare che i tratti caratteriali della persona siano dedotti dalla registrazione delle chiamate e dall'uso delle applicazioni; qui l'accuratezza assume valori non inferiori al 69%.

Ancora più curioso pensare che per captare lo stato d'animo degli utenti sia valutato il modo di digitare sulla tastiera; qui l'accuratezza è sempre superiore all'82%.¹⁶

¹⁶ Craked Labs CC BY-SA 4.0

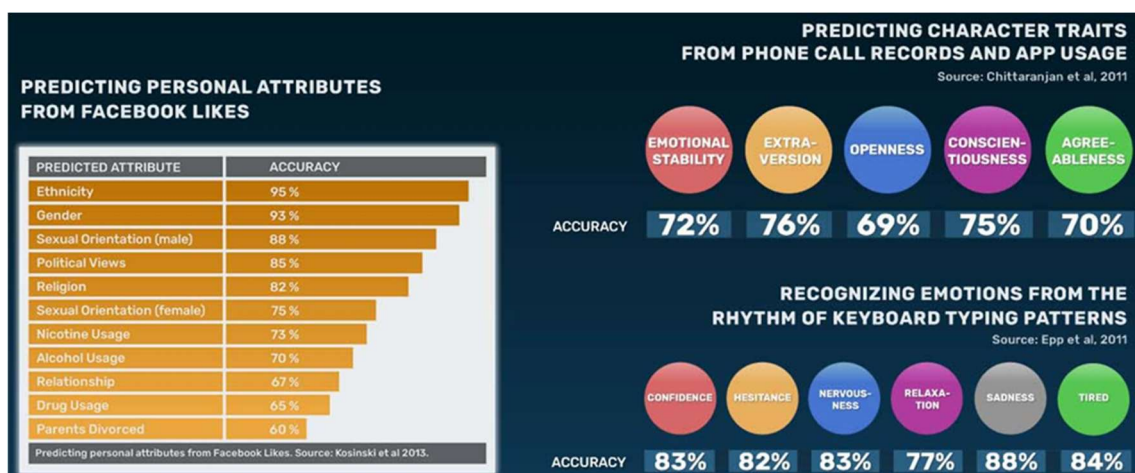


Figura 12: Le analisi tramite like - Corporate Surveillance in Everyday Life

Una delle ragioni per cui il tracciamento e la profilazione sono diventati così pervasivi è che quasi tutti i siti web, i fornitori di app e molti produttori di dispositivi condividono attivamente dati comportamentali con terze parti. Studi recenti hanno evidenziato che:

- Su un campione di un milione di siti web, sono stati individuati 80.000 servizi di tracciamento appartenenti a terze parti
- Un'analisi su 200.000 utenti tedeschi, che hanno visitato 21 milioni di pagine web, ha mostrato che i tracker di terze parti erano presenti nel 95% delle pagine.

È importante evidenziare come esista un mercato basato su elenchi di persone con nomi, indirizzi e altre informazioni personali. Ad esempio, il sito web dmdatabases.com ha offerto elenchi di email e indirizzi postali di utilizzatori di sedie a rotelle, persone che fanno uso di insulina o che soffrono di cancro al seno, depressione, impotenza o infezioni vaginali.

Particolare importanza hanno anche i dati raccolti da banche, compagnie assicurative e altri operatori finanziari. Questi, infatti, hanno accesso a informazioni estremamente sensibili sulle persone: saldi dei conti, interazioni con i servizi clienti, acquisti, modalità di guida, attività sanitarie. Tutte queste informazioni permettono, come vedremo in uno dei paragrafi seguenti di effettuare valutazioni creditizie e pricing del rischio sul cliente. Interessante notare come per effettuare queste valutazioni sia analizzata anche la grammatica e la punteggiatura dei messaggi di testo e lo stato della batteria del proprio telefono: chi scrive in modo più ordinato, coerente e stabile manifesta una maggiore affidabilità, costanza e capacità di pianificazione così come chi lascia spesso il telefono

scarico mostra una scarsa capacità di gestione e quindi una minore affidabilità. È opportuno specificare che non sono causalità reali ma correlazioni statistiche che sono utilizzate come proxy comportamentali.

A seguito di questa descrizione, si presentano tre campi principali all'interno dei quali questi dati sono utilizzati:

2.7.1 Marketing

I Big Data svolgono un ruolo fondamentale nel campo del marketing moderno: essi, infatti, rappresentano una miniera d'oro di informazioni che possono essere sfruttate dalle aziende per comprendere meglio i propri clienti e guidare in maniera più efficace le decisioni strategiche.

L'obiettivo principale è ottenere una panoramica completa dei consumatori, dei loro comportamenti di consumo, delle preferenze e delle esigenze, al fine di prendere decisioni più mirate e personalizzate. Con il loro utilizzo si vuole: trovare, attrarre e indirizzare nuovi clienti di valore, mantenere i clienti esistenti, massimizzare il loro valore a vita, prevenire il loro abbandono e, in generale, di influenzarli ad agire in determinati modi.

Ad essere analizzati sono:

- Parametri statici: caratteristiche della personalità degli utenti, preferenze comportamentali, preferenze di consumo, tendenze valoriali.
- Parametri dinamici: motivazioni della domanda, atteggiamenti, caratteristiche cognitive, consapevolezza del rischio, senso dell'esperienza, aspettative e soddisfazione

I principali obiettivi del marketing attraverso l'utilizzo dei Big Data sono:

- Retargeting: l'analisi dei dati permette di identificare segmenti di clientela e di comprenderne le caratteristiche demografiche, geografiche e comportamentali. Inoltre, consente di costruire profili cliente più dettagliati e di adattare le strategie di marketing in modo più mirato, identificando trend emergenti, nuovi modelli di consumo e nicchie di mercato ancora inesplorate. Spingendo così le aziende ad adattarsi all'offerta e a un aumento di competitività.

- Definizione dell'Ideal Customer Profile (ICP): i Big Data permettono di delineare con maggiore precisione il profilo del cliente ideale, valutando aspetti sociodemografici ma anche comportamentali. Questo è possibile, ad esempio, studiando la modalità di navigazione, i prodotti acquistati e la frequenza d'acquisto.
- ADV personalizzato: grazie all'analisi dei dati (il più delle volte generati dai cookie, file che memorizzano le interazioni dell'utente online), le aziende possono elaborare messaggi e offerte altamente personalizzate: script di telemarketing, materiali promozionali, contenuti online, annunci, sconti e prezzi variabili.
- I Big Data consentono di adattare le strategie promozionali in base alle preferenze individuali, aumentando così l'engagement e migliorando i tassi di conversione.
 - Nei negozi online, ad esempio, è possibile prevedere quanto un cliente potrebbe valere nel lungo periodo o quanto è disposto a pagare in quel determinato momento: ciò porta a differenziare i prezzi persino per lo stesso prodotto.
- Esperienze personalizzate: l'analisi dei dati permette alle aziende di comprendere meglio le esigenze individuali dei consumatori e offrire esperienze ad hoc: raccomandazioni di prodotti o servizi basati sugli acquisti precedenti, offerte esclusive o incentivi personalizzati. Questo approccio contribuisce a creare una relazione più solida tra l'azienda e il cliente, migliorando la fedeltà e l'esperienza complessiva del consumatore.

2.7.2 Valutazioni creditizie

Tradizionalmente, un rapporto di credito conteneva esclusivamente informazioni sulla storia dei pagamenti e dei debiti di un consumatore ed era fornito principalmente da banche, finanziatori, agenzie di recupero crediti e altre istituzioni. Oggi, con l'avvento del mercato dei Big Data, moltissimi altri dati sono utilizzati per questo genere di valutazioni. È importante sottolineare che l'utilizzo di tali dati per valutazioni creditizie differisca da paese a paese: ad esempio, in Belgio e Francia si utilizzano registri di credito pubblico, nel Regno Unito sono prese in considerazione informazioni extra, mentre in

paesi come Svezia e Finlandia i rapporti di credito possono essere utilizzati per scopi diversi dalla semplice valutazione creditizia.

Grazie ai Big Data, l'affidabilità creditizia è determinata considerando sia i comportamenti e le caratteristiche passate di un individuo, sia le relazioni con altre persone. Da questi dati si ricava un punteggio che classifica gli individui su una scala di affidabilità. Tuttavia, questi punteggi restano spesso opachi e possono produrre risultati arbitrari, diventando talvolta profezie che si autoavverano.

Numerose startup (ad esempio: Lenddo, Kreditech e Cignifi) utilizzano dati provenienti dai social media, dalle ricerche web e dai telefoni cellulari per calcolare l'affidabilità creditizia di una persona senza fare riferimento a informazioni legate a transazioni finanziarie. Alcune analizzano come un individuo compila un modulo online o naviga su un sito web, o come già anticipato, si osserva la grammatica e la punteggiatura dei propri messaggi e lo stato della batteria del telefono. Altri includono nel calcolo dei punteggi di credito anche i dati sugli amici di una persona su un social network; ad esempio, Facebook aveva registrato un brevetto per la valutazione del credito basata sui rating di credito degli amici di un utente.

Arrivati a questo punto non sorprendono grandi collaborazioni tra aziende come MasterCard, Telefonica (fornitore di reti mobili) e agenzie di reporting di credito come Experian.

Si vuole far notare però che, la capacità di questi dati di prevenire e prevedere sovraindebitamenti risulta particolarmente limitata e discutibile in quanto non è possibile prevedere molte delle principali cause di indebitamento: malattie, divorzi, perdita del lavoro o cattive condizioni di mercato.

Infine, va sottolineato che non esistono solo valutazioni di credito ma sono effettuate, utilizzando gli stessi dati, anche valutazioni del rischio per polizze assicurative, per il settore degli affitti o per la selezione degli inquilini, con l'obiettivo, ad esempio, di evitare affittuari problematici o morosi.

2.7.3 Informazioni mediche

I Big Data stanno trasformando il settore sanitario. Infatti, molti sistemi sanitari intelligenti e diverse applicazioni raccolgono e analizzano enormi quantità di dati. Essi provengono da cartelle cliniche e dispositivi di monitoraggio e servono a identificare i pazienti che traggono maggior beneficio da specifici interventi terapeutici o a prevedere l'evoluzione di malattie.

Oggi, i Big Data sono considerati uno dei settori tecnologici futuri più essenziali e promettenti e hanno attirato l'attenzione della comunità medica. Grazie ai Big Data, è possibile migliorare i risultati per i pazienti, personalizzare le cure, ottimizzare le relazioni tra paziente e operatore sanitario e ridurre i costi ospedalieri.

È importante far notare che molte aziende di dati e compagnie assicurative, collegate a questo settore, raccolgono informazioni sulla vita quotidiana dei consumatori per prevedere i loro rischi salutari.

Ciò che fa riflettere è che sono utilizzati dati demografici, struttura familiare, acquisti, proprietà di automobili al fine di prevedere rischi sanitari individuali come diabete, cancro, ipertensione o depressione, come accaduto realmente con la compagnia assicurativa Aviva. Un altro esempio riguarda la società di analisi sanitaria GNS Healthcare che calcola i rischi sanitari individuali dei pazienti a partire da cartelle cliniche, genomica e comportamenti dei consumatori, li classifica con dei punteggi e fornisce i risultati ad aziende assicurative.

La gestione dei Big Data in sanità presenta rischi e criticità soprattutto legati alla privacy e alla sicurezza dei pazienti. Questi rischi sono ancora più significativi con la diffusione del cloud computing per l'immagazzinamento delle informazioni. Sebbene il cloud offra un modello semplice e flessibile per estrarre dati, aumenta il rischio di divulgazione di informazioni sensibili; inoltre, più dati sono immagazzinati, maggiore è la necessità di larghezza di banda, con conseguenti aumenti dei costi e limitazioni nella velocità di elaborazione, che in casi estremi può risultare un fattore molto critico.

Infine, è fondamentale che queste informazioni, essendo estremamente sensibili, siano trattate con la massima attenzione e anche solo nel rispetto dei codici etici e delle norme sociali.

3 – Personal Data Brokers

All'interno di questo capitolo si vogliono analizzare le offerte proposte dai personal data broker, prendendo in considerazione aspetti legati alla trasparenza e ai servizi offerti, ponendo attenzione sugli elementi che accumulano queste aziende.

3.1 Acxiom

La prima azienda presa in esame è Acxiom (<https://www.acxiom.com/>).

Si nota immediatamente come all'interno del portale non siano presenti informazioni riguardo i prezzi proposti. Ciò che invece è messo immediatamente in evidenza è la loro value proposition: Acxiom promette di aiutare le aziende a generare il massimo valore possibile dai propri dati e dalla propria tecnologia. Questo obiettivo è perseguito tramite i seguenti principali target:

- Identificazione del 50-70% dei visitatori del sito anche senza l'uso dei cookie, un vantaggio rilevante considerate le attuali restrizioni normative e tecnologiche.
- Un tasso di conversione più che raddoppiato da visitatore a cliente.
- Un aumento del 154% negli acquisti da parte di clienti già esistenti; questo testimonia la forte attenzione alla fidelizzazione.
- Un incremento del 20% nella spesa media dei clienti più fedeli e ricorrenti.

Questa proposta di valore appare indubbiamente molto accattivante.

Un altro aspetto rilevante è la dichiarazione dei partner con cui Acxiom collabora, ben visibile già nella prima pagina del sito. Tra questi figurano Google, Adobe, Salesforce (piattaforma statunitense che supporta e aiuta le aziende nella gestione e nello sviluppo delle relazioni con i clienti), Treasure Data (piattaforma di gestione dei dati dei clienti) e Snowflake (data warehouse basato su cloud che supporta Amazon Web Services, Google Cloud Platform e Azure, la piattaforma cloud di Microsoft). Colpisce, seppur senza stupire, come tali collaborazioni rivelino una rete strettamente intrecciata tra i principali attori del settore tecnologico e dei servizi cloud. Non sorprende, dato l'elevato livello di integrazione e interdipendenza che caratterizza questo sistema, la presenza ricorrente delle stesse aziende partner.

Un dettaglio interessante è che la presenza dei partner è enfatizzata in modo costante e con formati simili anche in altri siti del settore (scritte scorrevoli al centro della pagina da

sinistra verso destra), il tutto a sottolineare la solidità della rete di collaborazioni e il livello di dettaglio raggiungibile grazie alle informazioni raccolte.

All'interno del sito sono offerti diversi prodotti e servizi, tra cui soluzioni di gestione dei dati dei clienti, analisi di marketing e soluzioni pensate ad personam. L'obiettivo dichiarato è quello di soddisfare, o meglio superare, le aspettative dei clienti finali, i quali ormai si aspettano che i brand conoscano le proprie esigenze meglio di quanto avvenga nella realtà.

A colpire particolarmente è la dichiarazione secondo cui Acxiom dispone di un database di più di 300 milioni di consumatori statunitensi digitalmente raggiungibili, corredati da decine di migliaia di attributi unici per consumatore: l'azienda afferma di poter raggiungere fino al 70% della popolazione digitale. Inoltre, nel sito è esplicitamente indicato che le aziende target principali appartengono ai settori già citati in precedenza: retail, finanza, salute, assicurazioni...

Un ulteriore elemento che emerge con chiarezza, come già sottolineato, è l'assenza di qualsiasi informazione sui prezzi: per ottenere un preventivo è necessario contattare un esperto tramite il sito. Tuttavia, già in questa fase preliminare, è richiesto di fornire numerosi dati personali. Il processo infatti prevede: un primo passaggio in cui è richiesto l'indirizzo mail e il paese di provenienza, un secondo in cui si deve inserire il proprio nome e il proprio cognome e infine è richiesto il titolo di studio, l'azienda in cui si lavora e il numero di telefono. Questa procedura consente ad Acxiom di costruire un patrimonio informativo molto ampio ancora prima di offrire concretamente un servizio; è evidente sin da subito l'approccio fortemente basato sulla raccolta dati.

Si intende specificare però che tramite il catalogo digitale AWS Marketplace gestito da Amazon Web Service sono state trovate alcune informazioni sui servizi offerti da Acxiom. Tra i primi nella lista c'è l'accesso a un dataset di segmentazione dei consumatori finlandesi per 11.050 \$ all'anno. Qui sono fornite informazioni socio-demografiche e di reddito, disponibili a un livello di dettaglio molto elevato (si parla di griglie di 250 m) tale da permettere segmentazioni e analisi a livello di quartiere. In particolare, è scritto che le famiglie sono classificate in 43 categorie, poi riunite in 11 gruppi. Ad essere prese in considerazione sono il reddito medio in valuta locale e la presenza di figli o di adulti

di età superiore ai 55 anni. È specificato inoltre che i dati sono stati resi anonimi e per ciascuna area geografica sono rappresentati non meno di 100 individui.

Un altro prodotto simile riguarda l'accesso, per 12.025 \$ l'anno, a un dataset che segmenta i consumatori di Hong Kong a livello di isolato, utile per analisi di marketing e pianificazione commerciale. Qui, oltre alle caratteristiche della precedente offerta, sono presi in considerazione anche la nazionalità e se l'immobile in cui risiede la famiglia è di proprietà o in affitto.

Ancora più dettagliato risulta il pacchetto dati India Urban PrimeMarkets, che fornisce informazioni per un'amplissima gamma di casi d'uso, fornendo dati sui consumatori per le 35 principali città metropolitane e sub metropolitane del Paese. Qui la suddivisione è fatta per codice PIN (l'equivalente del CAP italiano), con non meno di 40 nuclei familiari per area geografica. Per ciascun record sono presentati lo stato, il distretto, il numero totale di famiglie, la popolazione urbana e quella rurale, le famiglie urbane e quelle rurali, il reddito medio e la fascia di appartenenza. Il tutto è fornito a un prezzo di 25.500 \$.

Personalmente ritengo difficile valutare immediatamente se queste cifre siano alte o basse. Se si considera l'uso che spesso le aziende fanno di questi dati e i vantaggi che riescono ad ottenere, traducibili in fatturati di miliardi di euro, il prezzo sembra ragionevole; d'altro canto, vista la facilità con cui sono raccolti i dati (es. Acxiom che richiede già molti dati personali solo per parlare con un operatore) il prezzo potrebbe sembrare elevato.

In sintesi, Acxiom propone un'offerta che punta sul dettaglio e sulla completezza dei dati, supportata da una rete di partnership strategiche, mantenendo però un approccio poco trasparente sul piano dei costi fino, si pensa, al contatto diretto con un proprio esperto.

3.2 Circana

Il secondo sito preso in esame è quello di Circana (<https://www.circana.com/it-it>), azienda nata, come specificato nel capitolo 2, dalla fusione tra IRI e NPD Group. Circana si presenta come un player di riferimento nella raccolta, gestione e interpretazione dei dati relativi al comportamento dei consumatori.

Dal confronto con il portale di Acxiom emergono immediatamente due aspetti significativi.

In primo luogo, non sono fornite informazioni chiare e dirette riguardo ai prezzi: come per Acxiom, l'unica modalità per ottenere dati è quella di mettersi in contatto con un operatore. È interessante notare che, per poter procedere a tale contatto, l'utente deve rilasciare necessariamente le seguenti informazioni: nome, cognome, email, titolo di studio, numero di telefono, azienda di appartenenza, ruolo, dipartimento, settore e mercato di riferimento. In questo modo, Circana acquisisce un considerevole volume di dati sensibili solamente nella fase preliminare della richiesta di informazioni.

In secondo luogo, le partnership strategiche sono, anche qui, valorizzate e riportate con evidenza sul sito. Tra queste figurano aziende di primo piano come Google, Adobe, Experian, Meta e Comscore (azienda leader a livello mondiale nella misurazione di audience). Questo sottolinea l'importanza della condivisione dei dati e del valore aggiunto derivante dalla collaborazione tra aziende.

Secondo l'informativa sulla privacy, Circana raccoglie diverse tipologie di dati:

- Dati personali raccolti direttamente: dati di profilo e di registrazione, risposte ai sondaggi (molto importanti per l'acquisizione di informazioni demografiche a livello granulare (es. quartiere), informazioni su bevande e cibi consumati, livello di istruzione, pareri di prodotti e servizi, dati sui pagamenti effettuati e interazioni con i servizi Circana).
- Dati personali da terze parti: informazioni ricavate da database pubblici, partner commerciali e piattaforme di social media. È specificato che questi dati servono a comprendere come gli utenti interagiscono con i servizi di entrambi (Circana e il terzo). Si analizzano, ad esempio, cosa gli utenti guardano e poi acquistano e quali pubblicità risultano più efficaci.
- Dati personali raccolti automaticamente: tramite cookie e i pixel tag. Sono raccolti: indirizzi IP, informazioni sulla posizione, visualizzazioni di pagine (data e ora), tipo di dispositivo, sistema operativo, lingua e numerosi altri dati.

Questi dati sono utilizzati sia per adempiere ai contratti sia per i propri legittimi interessi aziendali, come: supporto alle operazioni di acquisizione e fusione, alla riorganizzazione

delle operazioni aziendali e alla personalizzazione dei servizi e delle offerte. Inoltre, questi dati (ad esempio quelli demografici: genere, dimensione del nucleo familiare, numero di figli...) possono essere divulgati a piattaforme e fornitori di rete di terze parti in forma aggregata o anonimizzata.

Circana fornisce alle aziende strumenti per capire dove e come le persone acquistano, andando a garantire accuratezza, precisione e un'elevata granularità dei dati.

Due esempi di servizio offerto, sotto forma di pacchetti, sono:

- Liquid Data Go: raccoglie dati provenienti dai punti di vendita (dove e quando un prodotto è stato venduto), dai panel dei consumatori (comportamenti tracciati mediante scontrini, carte di credito e app), dall'analisi delle vendite online e dalle politiche di prezzo nei negozi e li fornisce alle aziende clienti in un modo decisamente più strutturato. Questo servizio consente di segmentare gli acquirenti in base al comportamento, alla città e al rivenditore utilizzato, aiutando brand e inserzionisti a pianificare strategie più efficaci.

I risultati osservabili includono incremento notevole del ROI, aumento dei tassi di conversione e incremento nella spesa media. Il prezzo del servizio non è pubblicamente disponibile.

- PubTrack Digital: servizio di raccolta e analisi dati sul mercato degli e-book che, con dati provenienti da editori e distributori, fornisce un'immagine più chiara su dove determinati libri digitali sono venduti, bestseller e andamento del mercato, utili a pianificare la strategia digitale.

È stato trovato¹⁷ che per i soci di IBPA (Independent Book Publishers Association) è permesso l'accesso a questo servizio a tariffe ridotte:

- Per le aziende con un fatturato inferiore a 3.5 milioni di \$ è richiesta una cifra pari a circa 4000\$ l'anno più una tantum di 975\$
- Per le aziende con un fatturato inferiore ai 5 milioni di \$ è richiesta una cifra pari a circa 8000\$ con una tantum di 2000\$.

¹⁷ Circan PubTrack Digital – An IBPA Member Benefit – <https://www.ibpa-online.org/page/circanapubtrack>

- Le aziende con fatturato maggiore non accedono a queste tariffe scontate.

Si può concludere che, anche per Circana, le informazioni dirette sui prezzi sono difficili da reperire a differenza delle partnership strategiche e dei dati di segmentazione offerti. È evidente il ruolo centrale di queste aziende nella raccolta dei dati e nella successiva segmentazione di clienti e prodotti, sia per decisioni di marketing che per strategie di vendita.

Infine, non sembra esserci traccia di vendita diretta di database, ad esempio mailing list, riferite a singoli luoghi ma piuttosto a soluzioni pacchetto riferite a una determinata area di mercato.

3.3 Merkle

Merkle (<https://www.merkle.com/>) è un'azienda che si impegna a fornire soluzioni di livello mondiale per massimizzare il valore del portafoglio clienti dei propri committenti. È l'unica azienda riconosciuta da Forrester¹⁸ come presente in ogni fase del ciclo di vita del cliente. Si vuole sottolineare come, negli ultimi 30 anni, Merkle ha acquisito più di 36 altre aziende, tra cui fornitori di software e tecnologia, società di analisi dati e agenzie creative ed esperienziali, consolidando così, sempre più, la propria posizione di leader nel marketing e nella gestione dell'esperienza cliente. Quest'ultimo elemento rispetta la tendenza illustrata nel paragrafo 2.6 *Una tendenza del mercato: le acquisizioni*.

Sul sito aziendale sono messi in evidenza i partner aziendali, presentati con lo slogan: *We partner with the best*. Tra i quali figurano Adobe, Google, Salesforce, AWS (Amazon Web Services) e Figma (principale strumento di progettazione collaborativa).

Come già osservato per Acxiom e Circana, non sono presenti chiare indicazioni sui prezzi dei servizi e per ottenere maggiori informazioni è necessario compilare il solito form e rilasciare una serie di dati personali, con la prospettiva, ma non la certezza, di essere successivamente contattati da un operatore.

¹⁸ Forrester Research, Inc. è una delle principali società di ricerca e consulenza indipendente a livello globale

All'interno della propria informativa sulla privacy, Merkle elenca tutti i dati personali che potrebbero essere oggetto di elaborazione. Colpisce l'utilizzo del condizionale: vista la natura estremamente sensibile di queste informazioni, sarebbe auspicabile una comunicazione più trasparente e definitiva. Questo approccio, tuttavia, non sorprende se inserito nel contesto più ampio del settore, caratterizzato da politiche di prezzo poco chiare e da un'offerta commerciale che rimane di difficile comprensione senza un contatto diretto con l'azienda.

Tra i dati che potrebbero essere elaborati sono presenti:

- Dati di contatto: nome, cognome, indirizzo, numero di telefono ed email
- Dati demografici:
 - Riferiti all'individuo: sesso, età, professione, istruzione e lingue parlate
 - Nucleo familiare: numero di adulti, numero di bambini

Da notare che Merkle dichiara: *“We do not knowingly collect information from individuals under 18, but we may receive information about how many children are in your household and their year of birth or age range”* ma allo stesso tempo ammette di raccogliere dati sul numero di bambini presenti in famiglia e sulla loro età. È bene precisare che successivamente è scritto che farà “ogni sforzo” per escludere i minori di 18 anni dalle analisi. L'affermazione però, viste tutte le analisi sino ad ora condotte, rimane poco credibile.

- Dati commerciali e patrimoniali:
 - Informazioni su beni di proprietà personali dell'individuo e i relativi dettagli (data d'acquisto, valore dell'immobile, grandezza).
 - Informazioni geografiche: abitazione in zona cittadina o rurale.
 - Informazioni sui veicoli posseduti: valore, modello, se è di proprietà.
 - Informazioni sull'utilizzo dei devices elettronici: Tv, social media, Wi-Fi, app, accessi bancari.
 - Informazioni su acquisti effettuati.
 - Informazioni finanziarie: presenza di debiti, di investimenti.
- Dati relativi alla connessione a internet: indirizzo IP, device di connessione. Affermano anche di raccogliere informazioni sulle interazioni nella casella email
- Dati sensibili: patente, carta identità, numero di passaporto, geolocalizzazione precisa, dati genetici, carta di credito, orientazione politica, orientazione sessuale e orientazione religiosa.

Questa raccolta così granulare consente a Merkle di effettuare una segmentazione estremamente precisa dei consumatori. Non si tratta solo di distinguere i clienti in base ai parametri demografici generici ma creare profili dettagliati che combinano dati comportamentali, socio economici e persino sensibili.

Ciò che è venduto da Merkle sono dei pacchetti personalizzati in base alla richiesta specifica del cliente. Come nel caso di Circana, non è chiaramente visibile una vendita diretta di database, ma l'accesso a piattaforme e soluzioni di analisi, alcune delle quali disponibili tramite AWS Marketplace (così come accade per Acxiom).

Ad esempio:

- CXM On Demand: servizio che integra la tecnologia cloud nell'ecosistema Customer Experience Management (CEM) di un brand per gestire l'esperienza cliente in modo flessibile e in tempo reale. Sono utilizzati i dati in tempo reale per poter offrire al cliente esperienze preziose e naturalmente coerenti con il proprio profilo utente su tutti i canali. Non è indicato un prezzo preciso, ma è specificato che si tratta di un sistema pay-per-use.
- LoyaltyPlus: piattaforma per la fidelizzazione dei clienti per i principali marchi. Si tratta di un programma che integra la fidelizzazione nei programmi strategici dell'azienda. Si sottolinea come questo prodotto sia venduto attraverso il portale AWS, per cui, a differenza di quanto accadeva per Acxiom, non è indicato alcun prezzo pubblicamente. Per poterlo visualizzare si è tentato di accedere al sistema: è emerso che con l'utilizzo di una mail istituzionale la procedura risultava impossibile da completare, mentre con una mail personale era possibile avanzare più a fondo. Tuttavia, al momento della richiesta di dati sensibili come numero di carta di credito e codici bancari, l'operazione per ragioni facilmente deducibili, è stata interrotta. Questo episodio evidenzia ulteriormente la complessità e l'opacità dei processi di accesso ai servizi Merkle, che risultano poco trasparenti e difficilmente verificabili dall'esterno.

In conclusione, anche nel caso di Merkle emergono due tratti comuni alle altre aziende analizzate: da un lato la forte valorizzazione dei partner strategici, dall'altro l'elemento di opacità nella determinazione dei prezzi.

Inoltre, si nota come tutti e 3 i siti delle aziende siano molto orientati al mostrare i risultati finali raggiungibili piuttosto che a chiarire in modo trasparente le soluzioni effettivamente offerte e i relativi costi, questo per lo meno dal punto di vista pubblico.

Anche in questo caso non si rileva una vendita diretta di database cliente; tuttavia, in modo a tratti preoccupante, sono resi espliciti i molteplici dati personali che possono essere utilizzati per lo svolgimento delle attività aziendali.

Si intende precisare che rispetto all'elenco presente nel paragrafo 2.5 *I principali attori nel mercato*, il PDB Arvato non sarà preso in considerazione; tuttavia, l'analisi non si andrebbe a discostare in maniera chiara e significativa dalle tre precedenti. Pertanto, si è deciso di proseguire il capitolo con un esempio di PDB più recente e innovativo: Lenddo.

3.4 Lenddo

È bene specificare che Lenddo (<https://lenddoefl.com/>) non è un classico data broker in senso stretto, ovvero non vende dati finalizzati a scopi pubblicitari e di marketing, ma si occupa di fornire informazioni utili a supportare le decisioni creditizie per consumatori e per micro-PMI.

In questo senso, il modello di business di Lenddo si distingue dagli altri PDB analizzati, poiché non monetizza tramite il marketing o le campagne pubblicitarie personalizzate, ma attraverso degli scoring alternativi.

In particolare, Lenddo è in grado di effettuare un targeting basato su attributi, interessi e comportamenti dei consumatori, che può portare a un'esclusione, talvolta ingiustificata, dall'accesso a determinati servizi. Ad esempio, CNN Tech News ha riportato che alcuni prestatori utilizzano dati provenienti dai social network per determinare l'affidabilità creditizia dei potenziali clienti. Lenddo, nello specifico, terrebbe conto anche del comportamento finanziario degli amici Facebook di un richiedente (come eventuali ritardi nei rimborsi) per valutarne l'affidabilità complessiva.

Lenddo dichiara, all'interno del proprio sito, di utilizzare dati provenienti dai telefoni cellulari poiché dimostrano le capacità organizzative dell'utente. Tra i vari elementi considerati vi sono:

- Il modo in cui i contatti sono salvati nello smartphone: solo nome e numero di telefono oppure con l'email, l'azienda, l'indirizzo e la data di nascita. Avere informazioni più complete per Lenddo implica un maggiore grado di perfezionismo e organizzazione. Questi tratti possono essere rappresentati da persone con alto livello di consapevolezza e attenzione, che desiderano avere ordine e controllo su tutti gli eventi della loro vita e cercano di pianificare attentamente il loro futuro, dunque dei buoni clienti per una banca.
- La pianificazione presente sul calendario dello smartphone: consente di osservare il tempo trascorso in riunioni o in qualsiasi altro evento. Per Lenddo è importante osservare l'attenzione nella pianificazione delle attività future dell'individuo.
- La tipologia di applicazioni installate: ad esempio, per Lenddo una persona con molte app relative allo sport e alla salute, è considerata più attenta allo stile di vita e, quindi, potenzialmente più affidabile creditiziamente.

Oltre a questi dati, Lenddo effettua test psicometrici. Essi consistono in domande sulle preferenze in relazione a valori monetari e ritardi temporali, cercando di ottenere un profilo quantitativo dell'impulsività finanziaria e della probabilità di rimborso dei debiti dell'intervistato riassunta in un punteggio chiamato LenddoScore.

Per quest'analisi è utilizzato lo Score Confidence, che restituisce una Green flag se Lenddo ritiene il punteggio predittivo del rischio accurato, oppure un Red flag. Questo indice si basa su quattro indicatori chiave:

- Indipendenza: valutazione dell'autonomia nella risposta alle domande (Lenddo è in grado di capire se sono state fornite con l'aiuto di terzi o strumenti come l'intelligenza artificiale).
- Impegno: attenzione prestata dall'utente nella compilazione.
- Completamento: grado di completamento delle informazioni richieste.
- Errore nel processo: eventuali errori tecnici, come problemi di connessione.

Il sistema Score Confidence combina le informazioni dei moduli comportamentali, i dati demografici auto dichiarati, i metadata a disposizione (es. cookie, cronologia del browser,

transazioni commerciali) per restituire lo stato Green o Red. È stato statisticamente osservato che non concedere i prestiti agli utenti Red flag ha aumentato i benefici ottenuti dai clienti Lenddo.

Una volta raccolti tutti i dati, Lenddo calcola il LenddoScore, un punteggio di credito che varia da 0 a 1000 in cui i valori più alti rappresentano una minore probabilità di insolvenza. Secondo l'Università di Harvard, gli istituti finanziari che utilizzano LenddoScore hanno aumentato i tassi di approvazione del 50% e ridotto i tassi di insolvenza del 12%.¹⁹

Un altro vantaggio, ad esempio per le banche, è che utilizzando Lenddo, e quindi monitorando in tempo reale il comportamento dei consumatori, è possibile notare l'avvenire di eventi che potrebbero incidere sulla probabilità di insolvenza dell'individuo (es. malattie, divorzio), potendo così intervenire in maniera tempestiva.

A questo punto dell'elaborato, sembra quasi superfluo affermare come sul sito dell'azienda non ci siano informazioni dirette sui prezzi dei servizi forniti e, come sempre, si deve passare da un operatore lasciando, sin dal principio, la solita larga quantità di informazioni.

Di particolare rilievo è la collaborazione avviata con Google, che fornisce principalmente Google Cloud. Questo sistema fornisce server e potenza di calcolo scalabile, permettendo di analizzare miliardi di dati rapidamente. Inoltre, permette a Lenddo un risparmio di costi in quanto adotta un sistema di pagamento pay-per-use e non è costretto a comprare spazi sovradimensionati per gestire i periodi di picco.

Lo stesso sistema di pagamento è utilizzato dai clienti Lenddo, i quali andranno a pagare per il numero di valutazioni o verifiche effettuate. Questo consente a Lenddo di ottenere flussi di entrate più continui anche prima che siano disponibili tutti i dati necessari per la calibrazione dei modelli di previsione del rischio.

¹⁹ Lenddo: using data analytics to boost financial inclusion – Digital Innovation and transformation
Harvard University - <https://d3.harvard.edu/platform-digit/submission/lenddo-using-data-analytics-to-boost-financial-inclusion/>

Un esempio di soluzione lanciata da Lenddo è la eKYC (Know Your Customer) adottata per accelerare la verifica creditizia nelle Filippine a partire dal 2018. Con questa soluzione non è più richiesto un colloquio di persona o online per l'inserimento di nuovi clienti che utilizzano una carta o aprono un conto bancario. Lenddo dichiara di utilizzare l'intelligenza artificiale e l'analisi avanzata per riunire le migliori fonti di dati digitali e comportamentali e valutare accuratamente la disponibilità creditizia delle persone.

Lenddo raccoglie decine di migliaia di dati dai social media, tra cui Google, LinkedIn, Twitter e Facebook. Sul sito ufficiale è riportato un grafico che testimonia come l'azienda integra dati tradizionali (finanziari) con dati alternativi (sociali e psicometrici) per avere un quadro completo sul rischio di credito, fondamentali soprattutto per quei clienti con una scarsa storia creditizia.

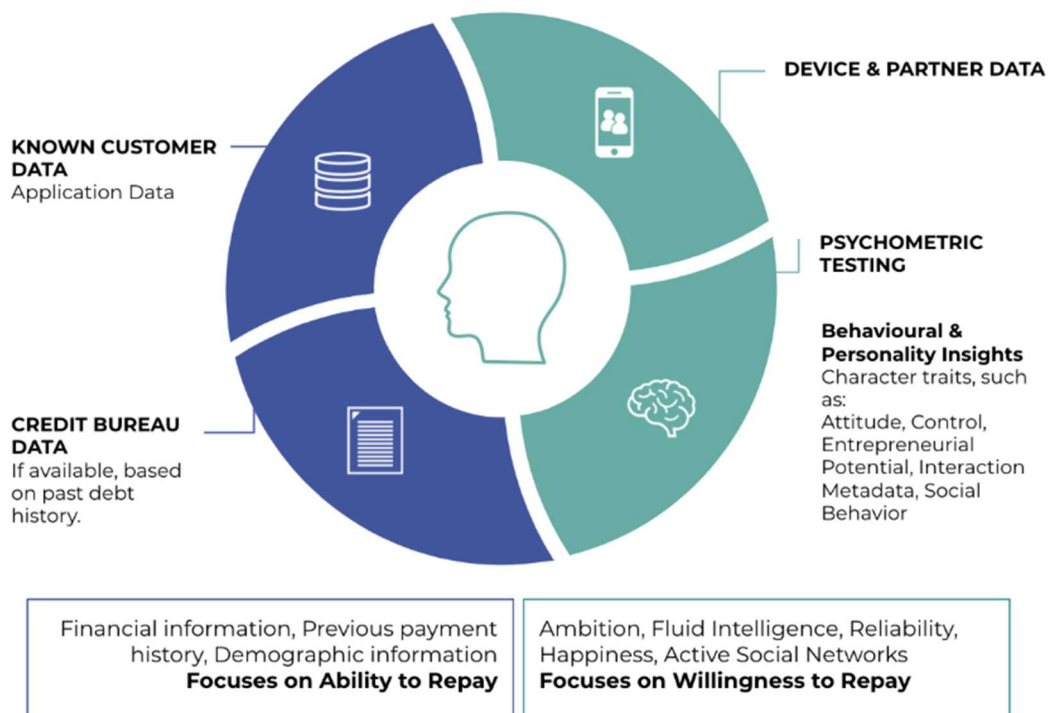


Figura 13: i dati di Lenddo, sito ufficiale

La parte sinistra del grafico si concentra sulla capacità di rimborso del cliente, è basata ovviamente su coloro che possiedono un minimo di storia creditizia. Qui sono raccolti i dati forniti dal cliente al momento della domanda di prestito e i propri dati storici finanziari.

La parte destra invece è rivolta a tutta quella fetta di clientela che non possiede una sufficiente storia creditizia. Tra i vari dati sono utilizzati quelli raccolti tramite i

dispositivi del cliente (telefono, computer o app), quelli provvisti dai partner esterni e quelli ottenuti dai test psicometrici.

Questa sezione di grafico è più rivolta a stabilire la propensione al rimborso del cliente.

In questo senso non stupisce il fatto che tra i primi stati in cui si è sviluppato il pacchetto eKYC sia stato al di fuori dell'Europa e proprio nelle Filippine; basta pensare che nel 2019 solo l'8% dei filippini era titolare di un conto di moneta elettronica, nel 2021 già passato al 36%. Così come i titolari di conti in banca sono aumentati dal 12% nel 2019 al 23% nel 2021. Numeri decisamente più piccoli rispetto a quelli europei.²⁰

Infine, al termine del paragrafo si vuole sottolineare che, come per le aziende citate in precedenza, non ci sono grandi informazioni numeriche per quanto riguarda i prezzi dei servizi offerti, per lo meno visibili pubblicamente.

A differenza delle altre aziende però, Lenddo non mette in risalto i partner strategici. A tal proposito si è venuti a conoscenza della relazione con Google non attraverso il sito di Lenddo ma tramite quello di Google.

Infine, è importante evidenziare che questi metodi di credit scoring alternativo sollevano questioni etiche rilevanti, poiché raccolgono informazioni dettagliate sui consumatori, inclusi comportamenti online e tratti psicologici, al fine di valutare l'affidabilità creditizia. In contesti poco regolamentati, la tutela della privacy e dei diritti dei consumatori assume un ruolo molto importante, soprattutto considerando relazioni come quella tra Lenddo e Google e il fatto che, come riportato nel paragrafo 2.6 *Una tendenza del mercato: le acquisizioni*, Google abbia acquisito la società israeliana di cybersecurity Wiz.

Al termine di questo capitolo, a supporto delle precedenti preoccupazioni, si vuole inserire un ultimo paragrafo riferito all'azienda americana DM Databases. Essa è specializzata nella creazione e successiva commercializzazione di database, anche

²⁰ Guida di settore – Settore bancario e finanziario, comprese le assicurazioni – Ambasciata d'Italia a Manila - https://ambmanila.esteri.it/wp-content/uploads/2024/04/Banking-Insurance-and-Finance-Sector-guide-draft_compressed-1.pdf

personalizzati, per poter contribuire all'ottimizzazione degli sforzi di marketing dei propri clienti.

3.5 DM Databases

L'azienda (<https://dmdatabases.com/>) si impegna nella commercializzazione di database, offrendo inoltre la possibilità di crearne uno su misura qualora le richieste del cliente non fossero soddisfatte già da quelli esistenti. A garanzia della qualità dei dati, sul sito è dichiarato che oltre il 50% dei record è verificato da almeno tre fonti, ed è inoltre prevista una garanzia di rimborso sull'accuratezza dei dati. Questo meccanismo tutela i clienti dal punto di vista della qualità delle informazioni e segnala la professionalità e la forte consapevolezza dell'azienda in questione. Nel sito, questa politica è indicata come un punto di distinzione rispetto ai concorrenti e, in base a quanto verificato, nessuna delle aziende precedentemente analizzate offre soluzioni simili.

Un punto di differenza rispetto ai precedenti paragrafi è che sono fornite più indicazioni, seppur non sempre chiarissime, sui servizi offerti e sui prezzi.

L'azienda, infatti, oltre a fornire elenchi email di consumatori con determinate caratteristiche, offre anche i seguenti servizi:

- *Consumer Email Append*: servizio volto ad aggiungere dati ai contatti dei consumatori già presenti. In particolare, integra dati demografici e psicografici, quali: età, reddito, patrimonio, valore della casa, etnia, numero di persone in famiglia, stato civile, tipo di auto e interessi vari (sport, passioni e ambiente). È possibile effettuare inoltre un'integrazione più approfondita tramite un'analisi dettagliata del profilo. L'accuratezza dei dati è garantita dalla politica di rimborso descritta precedentemente.

All'interno del portale è ripetuto più e più volte che l'integrazione è offerta a un prezzo notevolmente ridotto, variabile da 2 a 10 centesimi per indirizzo mail raggiunto. Tuttavia, per ottenere informazioni più precise su prezzi e servizi, è necessario compilare il consueto form, fornendo: nome, azienda in cui si lavora, numero di telefono (facoltativo), email, una breve descrizione del prodotto o

servizio da pubblicizzare, mercato di riferimento e dettagli geografici (ad esempio tutta l'Italia, una specifica regione o un quartiere).

Si intende specificare che si è provato a richiedere informazioni. Una volta inviati i primi dati, la prima risposta email è arrivata dopo qualche minuto dal direttore vendite, Ken Newton. Nella risposta è stato richiesto se si desiderava acquistare una lista di aziende o di consumatori e quali filtri applicare. Una volta specificata la richiesta di una consumer list una nuova email ha chiesto di dimostrare l'esistenza di un sito aziendale attivo correlato all'azienda. Questo era necessario per verificare l'effettiva esistenza dell'azienda e per assicurare la conformità rispetto al FTC SPAM Compliance Act e per prevenire l'utilizzo di indirizzi email non validi.

Per ovvie ragioni non si è potuto procedere. Si riscontra però una maggiore rapidità e dettaglio nelle risposte, rispetto alle altre aziende analizzate.

Curioso è il livello di dettaglio delle centinaia di classificazioni basate sullo stile di vita: sono disponibili centinaia di categorie, tra cui sport praticati o seguiti in TV, moda e accessori indossati (gioielli, prodotti di bellezza, abbigliamento), interessi culturali e artistici (teatro, cinema, fotografia), collezionismo, beneficenza, dieta seguita, letture e passioni per animali. Le altre classificazioni sono consultabili al link: <https://dmdatabases.com/databases/consumer-databases/lifestyle-marketing-lists/>

Le tariffe per l'integrazione delle email con fattori legati allo stile di vita sono:

1. Se si vuole acquistare solamente una lista di indirizzi postali di una categoria specifica, si parte da una spesa minima di 250\$.
2. Se si vuole anche il numero di telefono la spesa raddoppia: 500\$.
3. Se si vuole anche l'indirizzo email la spesa è quasi nuovamente raddoppiata: 975€.
4. Qualora ci si voglia concentrare più su una pubblicità mirata, ovvero che faccia comparire gli annunci pubblicitari alle persone giuste senza usufruire dei loro contatti diretti, è offerta ad una spesa minima di 1100\$

È decisamente sorprendente osservare tutte le possibili integrazioni che l'azienda in questione può generare partendo da un indirizzo email. È possibile, infatti, ricavare informazioni su affiliazioni politiche, destinazione di donazioni, proprietà immobiliari, dati finanziari, acquisti, condizioni di salute e veicoli posseduti.

Tra gli altri servizi offerti si trovano:

- *Consumer Email Append*: integrazione indirizzo email in base alle caratteristiche del consumatore.
- *Business Data Append*: analogo al primo servizio descritto ma rivolto alle aziende.
- *Business Email Append*: analogo al secondo servizio descritto ma rivolto alle aziende.
- *Email Verification*: verifica la validità degli indirizzi email forniti.
- *Reverse Email Append*: associa un indirizzo di posta elettronica a nome, indirizzo postale e altri dati.
- *Cell Phone Append*: analogo al precedente ma associa un numero di telefono.
- *Data Analytics*: analisi dei dati per supportare segmentazione, profilazione e modellazione predittiva.

Dal sito si è dedotto che tutti i dati sono raccolti tramite sondaggi, monitoraggio degli acquisti e cookies di navigazione. Si è notato che nel documento Privacy Policy²¹ è specificato che i dati presenti direttamente sul sito non sono condivisi con terzi; quelli ottenuti tramite altri metodi, invece, sono venduti a chi acquista le liste. Personalmente, ne consegue una difficoltà, forse alimentata dalla poca fiducia, nel distinguere quali dati provengano dal sito e quali da fonti esterne.

²¹ <https://dmdatabases.com/about-us/privacy-policy/>

All'interno del sito sono state osservate tutte le differenti liste di indirizzi di posta elettronica che sono vendute, queste riguardano sia i consumatori che le aziende. Di seguito si riportano le liste di consumatori più frequentemente acquistate (ordine alfabetico).

POPULAR CONSUMER LISTS

Absentee Owners Email List	Medicare Leads	Millennial Lists
Accredited Investor Leads	Musicians - Writers - Artists Email Lists	
Affluent People - High Net Worth Leads	Motorcycle Owner Lists	
Ailment Sufferers - Leads by Ailment Type	Newlyweds Marketing Lists	
Aircraft Owners List	Pre-Mover Leads - Homes For Sale	
Avid Reader Lists - Lists By Reading Group	New Mover Leads	
Boat Owner Leads	New Homeowner Leads	
Car Owner Leads - Automotive Lists	Nurse List - RN Leads	
CBD Users Lists	Occupant List	
Commercial Real Estate Investor Leads	People by Occupation List	
Consumers by Credit Score Leads	Pilot List - Leads	
Consumers by Hobbies & Interests Mailing Lists	Political Donors Email List	
Consumers by Lifestyle Interests	Puerto Rico Resident List	
Credit Score Marketing Lists	Recently Divorced List	
Diabetic Leads	Real Estate Investor Leads	
Eco Friendly Green Consumers List	Religious People Lists	
Electric Car Owners Marketing List	Renter Leads	
Equestrian – Horse Owners Email List	RV Owners List	
Ethnic Lists - People by Ethnicity	Soccer Fans Marketing Lists	
Fashionista Leads - Fashion Buyer Lists	Senior Lists	
Gambling Enthusiasts List	Solar Power Marketing Leads	
Golfer Lists	Stock Investors Marketing Leads	
Gun Owners Email List	Traveler Lists	
High School Students Email List	Turning 65 Leads	
Hispanic Lists - Hispanic Household Leads	Vacant Property Owners Email List	
Home Equity Leads - HELOC Leads	Vacation Property Owners Mailing List	
Homeowner Lists	Veteran Lists	
Households with Children List	Voters - Registered Voters by Party List	
Insurance Leads	Waterfront Home Owners Mailing List	
Luxury Buyer Leads	Yoga Enthusiasts Email List	

Si vuole far notare come all'interno di questo elenco siano presenti dati relativi a condizioni mediche (persone con specifiche malattie alimentari, diabetiche e persone iscritte a Medicare) e dati personali e identificativi (etnia, opinioni politiche e affiliazioni, stato civile, dettagli sulle proprietà immobiliari e, ad esempio, possesso di armi).

Analizzando nello specifico la categoria delle email relative ai diabetici, si osserva che il database contiene circa 12 milioni di indirizzi postali fisici e quasi 5 milioni di email (dato che non sorprende, poiché la maggior parte della popolazione diabetica possiede un indirizzo fisico e non necessariamente un'email, anche a causa dell'età).

È specificato che la lista ha un tasso di accuratezza del 90% ed è prevista anche qui la politica del rimborso gratuito qualora questa non sia rispettata. Inoltre, è timidamente indicato che per ottimizzare la schedulazione dei contatti sono utilizzati dati provenienti da Facebook, mentre le fonti principali rimangono diverse aziende di survey diabetiche.

Si notano inoltre essere riportate differenti categorie di diabetici si sono classificati e per cui sono riportati dei prezzi di vendita:

- Diabetici che hanno acquistato libri su stili di vita sani: 100\$ per mille contatti.
- Diabetici istruiti, appassionati di servizi su diabete, nutrizione ed esercizio: 150\$ per mille contatti
- Diabetici che hanno acquistato libri specifici (es. Glycemic, Load Diet, Diabetes Comfort Food Cookbook): 100\$ per mille contatti.

Importante notare che ciascuna categoria è caratterizzata da centinaia di migliaia, talvolta alcuni milioni, di indirizzi email. Dunque, sebbene i prezzi sopra citati sembrano ridotti rispetto ad altre aziende analizzate, è necessario rapportarli alla quantità e qualità dei dati disponibili.

Inoltre, è indicato che per acquistare una lista email è necessario un investimento minimo di 875\$.

Un altro esempio è rappresentato dalle Ethnic Lists – People by Ethnicity, per le quali non sono presenti informazioni di prezzo dettagliate ma è richiesta una spesa minima di 850\$. È probabile che, a seguito della compilazione del solito form e la verifica dell'effettività dell'azienda, siano forniti dettagli più precisi sui prezzi. Le liste coprono 185 gruppi etnici diversi, con alcune categorie particolarmente delicate dal punto di vista etico:

- Overt European Jewish Names: riferito a nomi e cognomi europei “chiaramente” identificabili come ebraici
- Particolarmente dettagliata è la classificazione degli afroamericani, suddivisa per:
 - Liste di afroamericani con cognomi olandesi
 - Liste di afroamericani con cognomi inglesi
 - Liste di afroamericani con cognomi francesi
 - Liste di afroamericani con cognomi irlandesi
 - Liste di afroamericani con cognomi islamici (qui è inserito anche un elemento religioso)

- Liste di afroamericani con cognomi scozzesi
- Liste di afroamericani con cognomi gallesi
- Unicamente afroamericani

Analogamente sensibile è il trattamento dei dati religiosi, considerata dati personali sensibili secondo le normative di protezione dei dati (GDPR incluso). Gli individui sono segmentati in base a fede e credo religioso aprendo a possibili rischi di discriminazione o uso improprio dei dati. Le classificazioni includono: buddisti, cattolici, ortodossi orientali, ortodossi greci, islamici-musulmani, ebrei, shintoisti²², mormoni, luterani e Sikh²³. Per questa sezione, così come per le liste etniche, non sono disponibili informazioni chiare sui prezzi.

In conclusione, si vuole affermare che all'interno del sito di DM Databases si riscontra un livello di dettaglio maggiore sia per quanto concerne la precisione dei dati offerti e sia per i prezzi, rispetto alle aziende analizzate precedentemente.

Come per le altre società, si pensa che informazioni più complete siano fornite solo attraverso conversazioni con operatori, con una maggiore rapidità nelle risposte.

Ciò che invece si discosta dalle quattro aziende precedenti è la scarsa trasparenza circa le fonti dei dati. Nelle precedenti era consuetudine che i partner di rilievo comparivano in primo piano nella pagina del sito, qui a comparire è solamente Facebook ed in maniera timida e occasionale. Inoltre, il documento sulla privacy è relativamente scarso rispetto a quella degli altri siti.

La scarsa chiarezza sulle origini dei dati potrebbe essere dovuta alla natura particolarmente sensibile delle informazioni commercializzate; tuttavia, una maggiore trasparenza potrebbe supportare le istituzioni nel controllo e nella maggiore regolamentazione.

²² Religione tipica del Giappone

²³ Religione nata nel Punjab (India)

Giunti a questo punto dell'elaborato si vuole analizzare il ruolo del dato non solo facendo riferimento ai personal data broker ma anche considerando l'impatto che essi possono avere a livello di mercato e concorrenza. In particolare, il capitolo successivo approfondirà un caso recente che evidenzia come la concentrazione e l'utilizzo dei dati possa influenzare fortemente le dinamiche competitive, generando rischi di abusi di posizione dominante e pratiche anticoncorrenziali.

Il caso antitrust in oggetto riguarda la multa di 2,95 miliardi di euro inflitta a Google da parte della commissione Europea per l'utilizzo della piattaforma pubblicitaria AdX.

4 – CASO ANTITRUST

All'interno di questo capitolo si vuole presentare ed analizzare uno dei più recenti casi antitrust che hanno a che fare, seppur indirettamente ma in maniera pur sempre decisiva, con il mercato dei dati.

Il caso si riferisce a un'indagine che la Commissione Europea aveva aperto nel 2021 e che riguarda il mercato della pubblicità online, dove il colosso del web avrebbe distorto la concorrenza favorendo i propri servizi a scapito di quelli dei fornitori concorrenti, degli inserzionisti e anche degli editori online. Una violazione che andrebbe avanti dal 2014.²⁴

4.1 Introduzione al caso

Il caso riguarda, come anticipato, la multa inflitta dalla Commissione Europea di 2,95 miliardi di euro a Google, il 5 settembre 2025, per pratiche abusive nella tecnologia pubblicitaria online.

La notizia ha subito un impatto notevole anche perché si tratta della seconda sanzione più alta di sempre nell'ambito della concorrenza (la stessa azienda aveva ricevuto nel 2018 una multa da 4 miliardi per il servizio Google Android).

4.2 Il caso

Prima di procedere con l'accusa promossa dalla Commissione Europea è bene chiarire alcuni fattori.

Google è una multinazionale tecnologica statunitense che basa la sua principale fonte di ricavo sulla pubblicità. In particolare, vende spazi pubblicitari sui propri siti e applicazioni e funge da intermediario tra inserzionisti, che vogliono collocare i loro annunci, ed editori (siti web e app di terzi) che mettono a disposizione quello spazio.

È importante specificare, sin dal principio, come Google prenda parte in tutti i passaggi della filiera AdTech. È così in grado di raccogliere, analizzare e memorizzare una enorme quantità di dati.

²⁴ Google Trump contro l'Ue – Marco Bresolin – La Stampa, pag. 10 – 06/09/2025

La filiera è organizzata come segue:

- I Publishers, ovvero gli editori, sono i siti web, giornali, applicazioni che offrono spazi pubblicitari.
- Gli Advertisers, gli inserzionisti, rappresentano quelle aziende che vogliono acquistare dello spazio pubblicitario.
- In mezzo ai due precedenti attori si colloca la Ad Exchanges, piattaforma di scambio pubblicitario. Un mercato digitale in cui domanda (inserzionisti) e offerta (spazi disponibili degli editori) si incontrano in tempo reale e permette la vendita degli spazi pubblicitari tramite aste.

Infine, per la gestione e collocazione in tempo reale di annunci pubblicitari non correlati a una ricerca, gli inserzionisti ed editori si affidano a:

- Publisher Ad Server: server utilizzati dagli editori per gestire gli spazi pubblicitari, ad esempio, aiutano a scegliere quale annuncio è meglio mostrare.
- Ad buying tool, strumenti di acquisto pubblicitario per inserzionisti che permettono di gestire automaticamente gli acquisti di spazi pubblicitari online, decidendo in tempo reale prezzo e collocazione.

La situazione di mercato precedentemente descritta è così rappresentata dalla Commissione Europea:

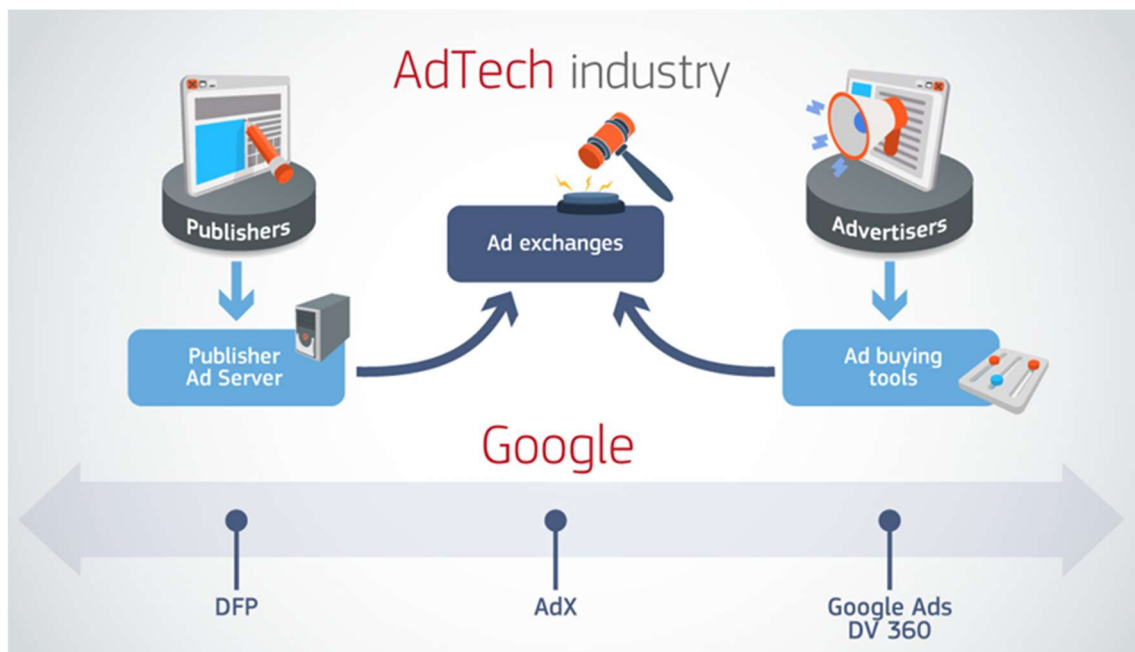


Figura 14: Commissione Europea

Dall'immagine appare, in maniera lampante, che Google è presente in tutti i segmenti della filiera. Questo espone l'azienda di pubblicità a potenziali conflitti di interesse, che, evidentemente, si sono tradotti in pratiche anticoncorrenziali.

Infatti, il ramo degli editori è rappresentato dal DFP (DoubleClick for Publishers), uno strumento di gestione degli annunci che consente agli editori di vendere, pianificare, pubblicare e gestire il proprio spazio pubblicitario; il ramo degli inserzionisti è caratterizzato da Google Ads (programma pubblicitario online tramite cui puoi creare annunci online per raggiungere gli utenti nel momento esatto in cui manifestano interessi per i prodotti che si offre ed è uno strumento che si può utilizzare per promuovere la propria attività commerciale) e DV 360 (elemento di Google che aiuta a lanciare campagne pubblicitarie digitali).

Infine, la piattaforma di scambio pubblicitario è AdX. Si tratta di un marketplace gestito da Google che consente a editori, inserzionisti, agenzie e reti pubblicitarie di scambiare impressioni in tempo reale tramite aste dirette.

4.3 L'accusa

Secondo la Commissione europea, Google ha favorito i propri servizi di tecnologia pubblicitaria a scapito di concorrenti, inserzionisti ed editori.

Inoltre, Teresa Ribera, Vicepresidente Esecutiva per una Transizione Pulita, Giusta e Competitiva della Commissione europea, sostiene che indirettamente ci sia stato un forte danneggiamento per i consumatori. Infatti, gli inserzionisti hanno dovuto sostenere costi di marketing più elevati, che probabilmente hanno scaricato sui consumatori europei sotto forma di prezzi più elevati. Le tattiche di Google hanno inoltre ridotto i ricavi degli editori, il che potrebbe aver portato a una qualità del servizio inferiore e a costi di abbonamento più elevati per i consumatori.

In particolare, la Commissione ha riscontrato che Google ha abusato della propria posizione dominante con due pratiche principali:

- Favoritismo verso AdX tramite DFP: Google informava in anticipo AdX della migliore offerta dei concorrenti in modo da poter offrire un rialzo minimo per vincere l'asta.

- Favoritismo verso AdX tramite Google Ads e DV360: strumenti progettati in modo da privilegiare sistematicamente AdX, evitando le piattaforme concorrenti. Rendendo così, questo mercato di scambio digitale molto più attraente.

Per queste ragioni Bruxelles ha inflitto a Google una multa di circa 3 miliardi di euro ordinandole di porre fine a queste pratiche di auto-preferenza e di attuare misure per eliminare i conflitti di interesse intrinseci lungo la catena adtech.

Google ha ora 60 giorni di tempo per informare la commissione su come intende procedere.

4.4 Il ruolo del dato

All'interno di questo caso Antitrust è centrale il ruolo del dato. Google, una delle più grandi aziende raccoglitrice di dati e detentrica di informazioni, essendo presente in tutti i rami della filiera ha la possibilità di intrecciare enormi quantità di dati, ricavando importanti informazioni e ottenendo così, un vantaggio strutturale nei confronti dei concorrenti.

In sostanza, conosce sempre le informazioni prima degli altri e per di più, le politiche da lei adottate spingono gli attori a utilizzare la propria piattaforma, alimentando una raccolta dati continua e pressoché inarrestabile, quasi a creare un circolo vizioso.

I dati raccolti sono sfruttati sia sul lato advertisers, orientando le transazioni verso AdX e sia sul lato degli editori, permettendo a Google di vincere sistematicamente le aste.

Inoltre, come anticipato, il fatto che la piattaforma di Google continui a essere utilizzata fa sì che siano raccolti sempre più dati, rafforzando in maniera netta il suo potere. Ciò le consente di offrire un targeting pubblicitario sempre più preciso e, di conseguenza, un match tra domanda e offerta pubblicitaria ancora più efficace, per il presente ma soprattutto per il futuro.

È chiaro che la concentrazione di dati in capo a un unico soggetto, in linea con la tendenza delle grandi aziende analizzata nel capitolo 2, ha reso possibile l'abuso di posizione dominante. Google, infatti, ha utilizzato i dati non solo per scopi pubblicitari ma anche per manipolare il mercato a proprio vantaggio.

4.5 Accenno di Teoria

In questa sezione si intende sottolineare come la classificazione di questo caso Antitrust come abuso di posizione dominante sia pienamente coerente con quanto studiato nei corsi. Infatti, si è in presenza di un mercato caratterizzato da una fortissima concentrazione in capo ad un'unica impresa, tale da configurarle una posizione di assoluto dominio.

È importante ricordare l'esistenza di due tipologie di abusi:

- Abusi di sfruttamento: sono prioritariamente condotti da un'impresa a danno dei consumatori finali
- Abusi di esclusioni: sono effettuati da un'impresa nei confronti delle imprese rivali con finalità esclusive.

Il caso analizzato rientra maggiormente all'interno della seconda categoria. Nello specifico si potrebbe parlare di un abuso di esclusione di tipo self-preferencing²⁵, in quanto Google ha sfruttato la propria posizione dominante in alcuni mercati della filiera pubblicitaria (derivanti principalmente dal proprio ruolo di leader nel mercato dei dati) per estendere il proprio potere all'interno delle piattaforme di intermediazione tra domanda e offerta pubblicitaria favorendo sistematicamente i propri prodotti o servizi rispetto a quello dei concorrenti.

4.6 Commento

Personalmente ritengo che queste sanzioni pecuniarie siano, seppur caratterizzate da cifre che possono sembrare spaventose, poco produttive e soprattutto tardive. Colpisce sempre il fatto che a colossi come Google sia concesso di compiere determinate operazioni, penso anche alle fusioni descritte nel capitolo 2, senza, a mio parere, un'adequata valutazione delle possibili conseguenze. Il fatto che Google, essendo forte in differenti mercati, cerchi di trarne vantaggio con pratiche più o meno trasparenti non mi sorprende. A lasciarmi spesso basito è l'inerzia delle istituzioni, che troppo spesso intervengono solamente dopo anni di lunghe indagini, quando i danni alla concorrenza sono ormai evidenti e realizzati, e i guadagni, in questo caso per Google, già consolidati.

²⁵ https://ec.europa.eu/commission/presscorner/detail/da/ip_25_1992

In rete è stata trovata una frase di Crider, professore onorario all'università UCL di Londra, che mi trova parzialmente d'accordo: *“Only a break-up will fix Google's monopoly, If Europe's enforcers flinch on a break-up in the end, Google will rightly chalk a fine up as a win”*. Egli sostiene come le multe, di fatto, siano considerate da Google come una vittoria: l'azienda non ha problemi finanziari né difficoltà ad affrontare lunghi e costosi processi giudiziari di ricorsi e controricorsi. Inoltre, in un mondo sempre più digitale, penso non sia così difficile assumere posizioni ricattanti.

La soluzione del professore, ovvero lo “spacchettamento” dell'azienda principale in tante piccole componenti autonome, penso possa essere efficace se gestita con molta attenzione e costantemente monitorata. Tuttavia, la considero difficilmente realizzabile, soprattutto in un contesto in cui operazioni come l'acquisizione di WhatsApp da parte di Facebook sono state autorizzate senza grandi esitazioni soprattutto quando accompagnate da dichiarazioni rassicuranti del tipo “manterremo i business completamente separati”.

Conclusione

All'interno dell'elaborato, partendo dalla definizione di dato e di Big Data, è stata evidenziata l'importanza crescente che essi rivestono nella società moderna.

I Big Data caratterizzati da volume, velocità, varietà, veridicità e valore, hanno assunto un ruolo sempre più centrale, reso possibile dallo sviluppo di diverse tecnologie tra cui: internet a banda larga sempre più stabile e veloce, la diffusione dei dispositivi dell'Internet of Things, l'intelligenza artificiale e il machine learning, e infine spazi di cloud computing sempre più ampi ed efficienti.

Queste tecnologie hanno fatto sì che la raccolta e l'analisi dei dati assumessero un ruolo sempre più cruciale all'interno dell'economia mondiale. A tal proposito il mercato dei Big Data è caratterizzato da una forte crescita negli ultimi anni, tanto che i dati sono stati definiti il “petrolio dell'era digitale”. In questo mercato, i principali attori sono i consumatori, che generano i dati, i content and service provider, che offrono i servizi digitali e raccolgono i dati prodotti dagli utenti, e i personal data broker, che utilizzano i dati degli utenti, fornendo loro ricompensa, per vendere i risultati delle loro analisi sul mercato.

È emerso chiaramente come il ruolo dei personal data broker, sebbene teoricamente fosse quello di redistribuire il valore dei dati verso i consumatori, sia spesso inefficace, sia sul piano concettuale che su quello matematico. Inoltre, il mercato mostra una tendenza alle acquisizioni in contrasto con l'obiettivo delle istituzioni di dare maggiore potere ai consumatori, sollevando interrogativi sulla reale efficacia di tali regolamentazioni.

La situazione è ancora più delicata quando i dati trattati riguardano informazioni sensibili, come quelle sanitarie e creditizie.

L'analisi dei personal data brokers ha evidenziato una scarsa trasparenza sui prezzi dei servizi e sui metodi applicati per il raggiungimento dei risultati. Si è notato inoltre una grande tendenza nel mostrare i principali partner strategici, quasi a funzionare da garanzia per chi vuole accedere a determinati servizi.

Di particolare rilievo è il caso Lenddo che utilizza, ad esempio, il salvataggio dei contatti telefonici, per fornire valutazioni creditizie sulle persone. È un caso a mio parere sorprendente soprattutto se si pensa che una persona potrebbe non ricevere un

finanziamento da una banca solamente perché ha salvato i contatti nella propria rubrica telefonica con nome e numero di telefono. Ancora più sorprendente è stato osservare i numeri di crescita nelle Filippine una volta che la soluzione Lenddo è entrata in azione. Infine, l'esame del caso Antitrust contro Google ha evidenziato che la concentrazione dei dati nelle mani di pochi grandi player possa portare ad abusi di posizione dominante. Inoltre, ha fatto emergere che sanzioni monetarie risultano spesso inefficaci di fronte a colossi globali, suggerendo che interventi preventivi sarebbero più utili, sebbene difficili da realizzare per la potenza economica di queste aziende e per la complessità geopolitica delle decisioni

Rimane tuttavia una certa perplessità riguardo alla possibilità di interventi incisivi in questo ambito, sia per la straordinaria potenza economica delle aziende coinvolte, che potrebbe comportare ripercussioni di natura nazionale (come dimostrato dalla reazione decisa del presidente degli Stati Uniti, Donald Trump, nei confronti dell'Europa nel caso analizzato), sia per la mostrata scarsa propensione delle autorità a intervenire anche di fronte a dichiarazioni quali “uniamo WhatsApp e Facebook ma manterremo i business separati”, agli occhi di molti palesemente irrealizzabile nella pratica sia per la natura delle aziende e sia per l'infinito valore che questi dati possono avere.

Si può dunque ritenere che il mercato dei Big Data abbia raggiunto un punto di non ritorno e che entrerà in maniera sempre più invasiva, consapevolmente o meno, nella vita di tutti i giorni di ciascun individuo.

Bibliografia e Sitografia:

- 📖 *Appunti per il corso: Sistemi Informativi Aziendali* – Luca Ardito, Fulvio Corno, Marco Torchiano – Politecnico di Torino – Dipartimento di Automatica e Informatica – 28/09/2022
- 📖 *Storia, definizione e ontologia dei dati* – Professoressa Rita Tegon – Rai Scuola
- 📖 *Data and Information Quality* – Carlo Batini, Monica Scannapieco – 2016 – Springer
- 📖 *New Horizons for a Data-Driven Economy. A roadmap for Usage and Exploitation of Big Data in Europe* – José María Cavanillas, Edward Curry, Wolfgang Wahlster – 2015 – Springer Open
- 📖 *Socioeconomic benefits of high-speed broadband availability and service adoption: A survey* – Wolfgang Briglauer, Jan Kramer, Nicole Palan – 2024
- 📖 *Big Data and Cloud Computing: Trends and Challenges* – Samir A. El-Seoud, Hosam F. El-Sofany, Mohamed Abdelfattah, Reham Mohamed – British University in Egypt -BUE, Cairo, Egypt – Novembre 2016
- 📖 *Il trattamento dei dati personali e strumenti di raccolta dei dati online: cookies e Big Data* – Arianna Fusaro – Università degli studi di Padova
- 📖 *The microeconomics of data – a survey* – Journal of Industrial and Business Economics – 06/07/2022
- 📖 *Corporate surveillance in everyday life* – Wolfie Christl – June 2017
- 📖 *The world's most valuable resource is no longer oil, but data* – David Parkins – The Economist – 14/10/2017
- 📖 *The microeconomics of data – a survey* – Flavio Pino – Journal of Business Economics – 06/06/2022
- 📖 *The Value of Personal Information in Online Markets with Endogenous Privacy* – Rodrigo Montes, Wilfried Sand-Zantman, Tommaso Valletti – Institute for Operation Research and Management Sciences located in Maryland, USA – 01/06/2018
- 📖 *Markets and Privacy* – Kenneth C. Laudon – Settembre 1996 – <https://dl.acm.org/doi/pdf/10.1145/234215.234476>
- 📖 *Do Consumers Benefit from Selling Their Data? The Economic Impact of Personal Data Brokers on Digital Markets* – Jan Kramer, Daniel Schnurr, Bastian Haberer – Bavarian State Ministry of Science and Arts

- 📌 *Rapporto 2014: Il Grand Tour del XXI secolo: l'Italia e i suoi territori* – il Mulino
- 📌 *Pros and Cons of Big Data* – Harvard on Digital – 28/02/2024 –
<https://harvardonline.harvard.edu/blog/pros-cons-big-data>
- 📌 *Big data: The next frontier for innovation, competition and productivity* – James Manyka, Michael Chui, Brad Brown, Jacques Bughin, Richard Dobbs, Charles Roxburgh, Angel Hung Byers – McKinsey Global Institute – Maggio 2011 –
https://www.mckinsey.com/~/media/mckinsey/business%20functions/mckinsey%20digital/our%20insights/big%20data%20the%20next%20frontier%20for%20innovation/mgi_big_data_exec_summary.pdf
- 📌 *Big data and data analytics. The potential for innovation and growth* – European Parliament – Settembre 2016 –
[https://www.europarl.europa.eu/RegData/etudes/BRIE/2016/589801/EPRS_BRI\(2016\)589801_EN.pdf](https://www.europarl.europa.eu/RegData/etudes/BRIE/2016/589801/EPRS_BRI(2016)589801_EN.pdf)
- 📌 *Big Data & Advanced Analytics per il Risk Management* – AIFIRM – Aprile 2022
<https://www.aifirm.it/wp-content/uploads/2016/03/2022-Position-Paper-35-Big-Data.pdf>
- 📌 *8 steps in the data life cycle* – Tim Stobierski – Harvard Business School – 02/02/2021
<https://online.hbs.edu/blog/post/data-life-cycle>
- 📌 *Introduzione ai Big Data ed esempi applicativi* – Marco Bianchi – Fondazione Ugo Bordoni – 21/09/2023
https://www.fub.it/wp-content/uploads/2023/09/Presentazione_21.09.2023.pdf
- 📌 *Qualità dei dati* – Docs Italia
[https://docs.italia.it/AgID/documenti-in-consultazione/lg-opendata-docs/it/bozza/aspetti-organizzativi-e-qualit%C3%A0-dei-dati/qualit%C3%A0-dei-dati.html#:~:text=SI%20RACCOMANDA%20di%20garantire%2C%20per,e%20attualit%C3%A0\)%2C%20come%20da%20indicazioni](https://docs.italia.it/AgID/documenti-in-consultazione/lg-opendata-docs/it/bozza/aspetti-organizzativi-e-qualit%C3%A0-dei-dati/qualit%C3%A0-dei-dati.html#:~:text=SI%20RACCOMANDA%20di%20garantire%2C%20per,e%20attualit%C3%A0)%2C%20come%20da%20indicazioni)
- 📌 *SISTEMI INFORMATIVI Definizioni, classificazioni*
<http://static.gest.unipd.it/labtesi/eb-didattica/GIA/sistemi-informativi.pdf>
- 📌 *Cosa sono i Big Data?* – Michael Chen – Content Strategist – 23/09/2024
<https://www.oracle.com/it/big-data/what-is-big-data/>
- 📌 *Big Data: definizione, benefici e sfide (infografica)* – Parlamento europeo – 23/03/2023

- <https://www.europarl.europa.eu/topics/it/article/20210211STO97614/big-data-definizione-benefici-e-sfide-infografica>
- 📌 *Big Data e tradizioni a confronto*
<https://www.purestorage.com/it/knowledge/big-data/big-data-vs-traditional-data.html#:~:text=Veridicit%C3%A0%3A%20i%20Big%20Data%20non,sulla%20qualit%C3%A0%20dei%20Big%20Data.>
- 📌 *Dati strutturati e dati non strutturati: qual è la differenza?* – Alexandra Jonker, Alice Gomstyn – IBM – 07/02/2025
<https://www.ibm.com/it-it/think/topics/structured-vs-unstructured-data>
- 📌 *5 V dei Big Data, cosa sono, quale ruolo rivestono* – Mirella Castigli – 28/08/2023
<https://www.zerounoweb.it/big-data/5-v-dei-big-data-cosa-sono-quale-ruolo-rivestono/>
- 📌 *LE TRE V DEI BIG DATA* – Alessandro Rezzani – 10/01/2018
<https://www.dataskills.it/le-tre-v-dei-big-data/#grefu>
- 📌 *Qualità dei dati* – Domenico Natale – La comunicazione – Agosto 2025
<https://www.lacomunicazione.it/voce/qualita-dei-dati/>
- 📌 *Data lifecycle: The 8 staged and who is involved* – Gareth Davies – KNIME – 23/08/2024
<https://www.knime.com/blog/the-data-lifecycle>
- 📌 *A Short History of Big Data* – Big Data Framework Alliance – Enterprise Big Data framework – 18/03/202
<https://www.bigdataframework.org/knowledge/a-short-history-of-big-data/>
- 📌 *Big Data: Cosa Sono, la Storia, le Caratteristiche, le Analisi ed Esempi* – Andrea Magnani – Marketing Freaks
<https://www.themarketingfreaks.com/2019/11/big-data-cosa-sono-la-storia-le-caratteristiche-le-analisi-ed-esempi/>
- 📌 *The History and Evolution of Big Data* – Extentia – 30/10/2024
<https://www.extentia.com/the-history-and-evolution-of-big-data/>
- 📌 *Big Data e l'approccio secondo il Censo Romano* – William Inmon – 07/02/2014
<https://www.itware.com/blog-itware/big-data-analytics-data-management/item/1012-big-data-e-lapproccio-secondo-il-censo-romano>
- 📌 *Rivoluzione digitale: 5G e Big Data trasformano l'esperienza del cliente* – Felipe Ferreira – 05/09/2024

- <https://tiinside.com.br/it/05/09/2024/revolucao-digital-5g-e-big-data-transformando-a-experiencia-dos-clientes/>
-  *Descubra o impacto do 5G no monitoramento de mídia: coleta e análise de dados em tempo real* – info4 – 28/05/2024
<https://blog.info4.com.br/2024/05/28/descubra-o-impacto-do-5g-no-monitoramento-de-midia-coleta-e-analise-de-dados-em-tempo-real/>
 -  *Big Data & Machine Learning (How do they relate?)* – WEKA – 12/10/2021
<https://www.weka.io/learn/glossary/ai-ml/big-data-machine-learning/>
 -  *Big Data & Cloud Computing: The Roles & Relationships* – Gaurav Sharma – IEE Computer Society – 02/18/2020
<https://www.computer.org/publications/tech-news/trends/big-data-and-cloud-computing>
 -  *Nasscom, Crisil launch 'Big Data-The Next Big Thing' report* – 05/09/2012 -
<https://www.moneycontrol.com/news/business/stocks/-1887295.html>
 -  *Big Data Market Summary* – Aarti Dhaptel – Market Research Future – Luglio 2025 – <https://www.marketresearchfuture.com/reports/big-data-market-7846>
 -  *Big Data Market Size and Share Outlook - Forecast Tends and Growth Analyst Report* – Ketan Gandhi, Piyush Gautam – EMR Claight – 2024 –
<https://www.expertmarketresearch.com/reports/big-data-market>
 -  *Big Data Market* – MarketsandMarkets™ – Gennaio 2024 –
<https://www.marketsandmarkets.com/Market-Reports/big-data-market-1068.html>
 -  *Big Data Market – Industry Analysis and Forecast (2025 – 2032) Trends, Statistics, Dynamics* – Stellar Market Industry -
<https://www.stellarmr.com/report/Big-Data-Market/1141>
 -  *Il mercato italiano dei Big Data vale 3,42 miliardi di euro, +20%* – Polimi School of Management – 2024 – <https://www.osservatori.net/comunicato/big-data-business-analytics/big-data-italia-mercato/>
<https://www.osservatori.net/comunicato/big-data-business-analytics/mercato-big-data-crescita/>
 -  *Big data, mercato italiano vale 3,42 miliardi di euro (+20%)* – la Repubblica – 05/11/2024 –
https://finanza.repubblica.it/News/2024/11/05/big_data_%C2%A0mercato_italia_no_vale_3_42_miliardi_di_euro_+20percento_-71/

-  *Big Data, è boom per il mercato italiano: +20%* – Alessandro Piva, Carlo Vercellis – Politecnico di Milano – 06/11/2024 – <https://www.zerounoweb.it/big-data/big-data-e-boom-per-il-mercato-italiano-20/>
<https://www.industriaitaliana.it/mercato-big-data-italia-2024-politecnico-di-milano-il-mercato-dei-dati/>
-  *Come guadagnare con Google Opinion Rewards* – Alessandro Nodari – smartworld – 11/03/2024 – <https://www.smartworld.it/guide/google-opinion-rewards-come-guadagnare.html>
-  *Ricevere premi – Guida di Google Opinion Rewards* – <https://support.google.com/opinionrewards/answer/7378183?hl=it#zippy=%2Cquanto-viene-pagata-la-partecipazione-a-ogni-sondaggio%2Cperch%C3%A9-non-ho-ricevuto-il-pagamento-per-un-sondaggio%2Cesiste-un-modo-per-rispondere-a-un-maggior-numero-di-sondaggi-per-guadagnare-di-pi%C3%B9%2Ccome-faccio-a-trasferire-i-miei-pagamenti-da-google-opinion-rewards-a-paypal>
-  *Technology & Data at Bertelsmann* – Bertelsmann – Markus Harbaum – <https://www.bertelsmann.com/company/alliances/technology-and-data/#st-1>
-  *Mergers and Acquisitions* – M&A Equilibrium – 31/01/2025 – <https://mandaequilibrium.com/mergers-and-acquisitions-in-big-data/>
-  *Nbc-Universal a Comcast per 13,75 mld usd, nasce colosso mondiale* – L'eco di Bergamo – 03/12/2009 – https://www.ecodibergamo.it/stories/apcom/nbc-universal-comcast-1375-mld-usd-nasce-colosso-mondiale-o_104799_11/
-  *Treccani – Dizionario di Economia e Finanza (2012)*
[https://www.treccani.it/enciclopedia/comcast_\(Dizionario-di-Economia-e-Finanza\)/](https://www.treccani.it/enciclopedia/comcast_(Dizionario-di-Economia-e-Finanza)/)
[https://www.treccani.it/enciclopedia/tesco_\(Dizionario-di-Economia-e-Finanza\)/](https://www.treccani.it/enciclopedia/tesco_(Dizionario-di-Economia-e-Finanza)/)
-  *Cross-Device Exit: Telenor Acquires Tapad For \$360 Milion* – Allison Schiff – adexchanger – 01/02/2016
<https://www.adexchanger.com/mobile/cross-device-exit-telenor-acquires-tapad-360-million/>
-  *I giganti tecnologici: ecco le dieci operazioni più importanti di sempre* – Biagio Simonetta – Il Sole 24 Ore – 14/06/2016 – [Giugno 2016 - Microsoft acquista LinkedIn per 26,2 miliardi di dollari](#)

-  *Così LinkedIn può ridisegnare il futuro di Microsoft. Ecco perché Nadella lo ha voluto a tutti i costi* - Biagio Simonetta – Il Sole 24 Ore – 13/06/2016 – [Così LinkedIn può ridisegnare il futuro di Microsoft. Ecco perché Nadella lo ha voluto a tutti i costi - Il Sole 24 ORE](#)
-  *Facebook-WhatsApp, la paura degli utenti: “La nostra privacy è a rischio, boicottiamoli”* – Katia Riccardi – la Repubblica – 21/02/2014 – https://www.repubblica.it/tecnologia/2014/02/21/news/facebook_WhatApp_fusione_paure-79271799/
-  *Microsoft compra LinkedIn* – Diletta Parlangeli – WIRED – 16/06/2016 – <https://www.wired.it/economia/finanza/2016/06/13/microsoft-compra-linkedin-2/>
-  *Google acquista l'israeliana Wiz: un accordo tra luci e ombre. Non solo un caso economico* – Laura Aprati – RaiNews.it – 30/03/2025 – [Google acquista l'israeliana Wiz: un accordo tra luci e ombre. Non solo un caso economico](#)
-  *Big data nel marketing: comprendere i consumatori e personalizzare le strategie* – EPICODE Institute of Technology – <https://epicode.com/it/big-data-nel-marketing-comprendere-i-consumatori-e-personalizzare-le-strategie/>
-  *Big data: definizione, a cosa servono ed esempi per il marketing* – Luca Bizzarri – <https://blog.ofg.it/big-data/>
-  *Big Data Usage in Marketing Research* – Shi Zheng, Xia Jin, Wen Zheng – School of Economic and Management Liaoning (China), Northeastern University in Qinhuangdao (China) – 2022
-  *Problems of Big Data Adoption in the Healthcare industries* – Surya Kant Pal – Sharda University – Dicembre 2021 – https://www.researchgate.net/publication/357024187_Problems_of_Big_Data_Adoption_in_the_Healthcare_Industries
-  *Google hit with \$3.5B fine from European Union in ad-tech antitrust case* – The Associated Press – ABC News – 05/09/2025 – <https://abcnews.go.com/Technology/wireStory/google-hit-35-billion-fine-european-union-ad-125295028>
-  *Google fined €2.95 bn by EU for abusing advertising dominance* – Liv McMahon – BBC – 05/09/2025 – <https://www.bbc.com/news/articles/c1wgn3lre14o>
-  *EU fines Google nearly €3bn for ‘abusing’ dominant position in ad tech* – The Guardian – 05/09/2025 –

- <https://www.theguardian.com/technology/2025/sep/05/google-fined-european-union>
- ✚ *Google hit with \$3.5 billion fine from European Union in ad-tech antitrust case* – Kelvin Chan – FOX 5 News – 05/09/2025 – <https://fox5sandiego.com/news/ap-google-hit-with-3-5-billion-fine-from-european-union-in-ad-tech-antitrust-case/>
 - ✚ *Ue: multa di 2,95 miliardi a Google, distorta concorrenza. Azienda annuncia ricorso* – Rai News.it – 05/09/2025 – <https://www.rainews.it/articoli/2025/09/ue-multa-google-distorta-concorrenzaazienda-annuncia-ricorso-acd67f7d-3fd4-4cee-b579-1ca48dca767c.html>
 - ✚ *Commission fines Google €2.95 billion over abusive practices in online advertising technology* – European Commission – Press release – 05/09/2025
https://ec.europa.eu/commission/presscorner/detail/en/ip_25_1992
https://ec.europa.eu/commission/presscorner/api/files/document/print/en/ip_25_1992/IP_25_1992_EN.pdf
 - ✚ *Google multa da UE per quasi 3 miliardi per distorta concorrenza. La società: faremo ricorso* – la Repubblica – 05/09/2025 –
https://finanza.repubblica.it/News/2025/09/05/google_multa_da_ue_per_quasi_3_miliardi_per_distorta_concorrenza_la_societa_faremo_ricorso-211/
 - ✚ *Consumer data and the digital economy, Emerging issues in data collection, use and sharing* – Phuong Nguyen & Lauren Solomon – Consumer Policy Research Centre – https://cprc.org.au/wp-content/uploads/2021/12/Full_Data_Report_A4_FIN-2.pdf
 - ✚ *LenddoEFL Lancia la soluzione eKYC per accelerare la verifica nelle Filippine* – LendoEFL – 27/09/2018 – <https://lenddoefl.com/news/2018/8/27/press-release-lenddoefl-launches-ekyc-solution-to-speed-up-verification-in-the-philippines>
 - ✚ *LendoEFL: servire meglio i clienti con algoritmi basati sui dati ospitati su Google Cloud* – Google Cloud and LendoEFL – https://cloud.google.com/customers/lenddoefl/?utm_source
 - ✚ *La Personalità Può Predire L'andamento Del Prestito?* – Di Mondato – LendEFL – <https://lenddoefl.com/news/2021/4/11/mondato-feature-can-personality-predict-loan-performance>
 - ✚ *Lenddo: using fata analytics to boost financial inclusion* – Digital innovation and transformation Harvard University – 09/04/2018 –

<https://d3.harvard.edu/platform-digit/submission/lenddo-using-data-analytics-to-boost-financial-inclusion/>

-  *Sito ufficiale Acxiom:* <https://www.acxiom.com/>
-  *Sito ufficiale Circana:* <https://www.circana.com/it-it>
-  *Sito ufficiale Merkle:* <https://www.merkle.com/>
-  *Sito ufficiale Lenddo:* <https://lenddoefl.com/>
-  *Sito Ufficiale DM Databases:* <https://dmdatabases.com/>