

POLITECNICO DI TORINO

Corso di Laurea Magistrale in Ingegneria Gestionale



DATA BROKER: IMPATTO ECONOMICO E CONSEGUENZE SULLA PRIVACY

Relatore

Prof. Carlo Cambini

Candidato

Elia Corrado Campione

Anno Accademico 2023/2024

Indice

SEZIONE 1- LETTERATURA	4
1.1 DEFINIZIONE DI DATA BROKER	4
1.2 LE FONTI DEI DATI DEI DB.....	5
1.3 IMPORTANZA DEI DB.....	7
1.4 TIPOLOGIA DI DB	8
1.5 DATI E SERVIZI SU CUII DB FONDANO IL LORO BUSINESS.....	11
1.6 CHI SONO I PRINCIPALI DB	13
1.7 MODELLI SUL MERCATO DEI DB: analisi della letteratura economica	27
1.8 RISCHI LEGATI ALLA FIGURA DEI DB.....	46
SEZIONE 2 – POLICY	49
2.1 IL DATA ACT.....	49
2.2 LA REGOLAZIONE DEI DB	51
2.3 IL GDPR.....	52
2.4 DB E GDPR	56
2.5 IL CAPITALISMO DI SORVEGLIANZA	63
2.6 IL PROCESSO DI “COMMODIFICATION”	65
2.7 POLICY PROPOSTE DALLA LETTERATURA ECONOMICA	67
2.8 I CASI KOCHAVA E X-MODE.....	72
2.9 DB E GEOPOLITICA, CASO EQUIFAX E TIKTOK.....	76
SEZIONE 3 – DB, PREZZI, CLIENTI, CONSUMATORI E DATASET	78
3.1 STRUTTURE DI PREZZO NEL MERCATO DEI DATI	78
3.2 STRUTTURE DI PREZZO UTILIZZATE DAI DB	81
3.3 DB E CLIENTI	82
3.4 DB E CONSUMATORI	83
3.5 I DATASET DEI DATA BROKER.....	89
3.6 GEOGRAFIA DEI DATASET	97
3.7 SOVRAPPOSIZIONI E CONFRONTO TRA DATASET	98
3.8 LA RISPOSTA DEL MERCATO AI DB	105
3.9 CONCLUSIONI.....	108
Bibliografia e sitografia	109

INTRODUZIONE

Dagli ultimi dieci anni, dispositivi elettronici, siti web e app utilizzano tecniche sempre più sofisticate per estrarre dati di qualsiasi tipo dai consumatori. Inoltre, lo storage dei dati è diventato più conveniente, incentivando le aziende a conservare dati anche quando non più necessari per gli scopi originali. Di conseguenza, molte piattaforme digitali hanno basato i loro modelli di business sulla condivisione di dettagli degli utenti con inserzionisti, aziende private e agenzie governative, spesso tramite intermediari noti come data broker.

I data broker sono diventati attori chiave delle tecnologie basate sui dati. La loro attività consiste prevalentemente nel raccogliere e rivendere ingenti quantità di dati sugli individui, incluse informazioni sensibili su dati demografici, finanziari, sanitari, interessi e acquisti. I data broker si affidano a una vasta gamma di fonti, come social media, servizi di ricerca, app, programmi di fedeltà, fornitori di pagamenti con carta e registri pubblici. La copertura e la profondità dei loro dataset sono senza pari. Diversamente dalle piattaforme che godono di notorietà, i data broker non interagiscono direttamente con i consumatori, che spesso non sono consapevoli della loro esistenza. Tuttavia, molti aspetti della vita quotidiana, dalla pubblicità online alla determinazione dei prezzi da parte delle aziende, dal punteggio di credito alla gestione del rischio, sono mediati da questi attori. A tal proposito, sul sito di Avast si legge: "You may not know who data brokers are, but they know you".

Principalmente a causa dell'opacità delle loro operazioni, il comportamento dei data broker rimane in gran parte inesplorato. Questa industria, tanto oscura quanto pervasiva, nel 2018 generava ricavi di oltre 200 miliardi di dollari l'anno. Il settore è dominato da poche grandi aziende, quali Oracle, Acxiom, Experian, sostanzialmente non regolamentate dal punto di vista della concorrenza. L'Unione Europea ha introdotto nuove normative per i mercati digitali, come il Digital Service Act, il Digital Markets Act e il Data Act, ma il loro effetto sul mercato dei data broker è ancora limitato.

Recentemente, le autorità antitrust hanno iniziato a indagare sui data broker e sulle loro pratiche di vendita dei dati. Esempi includono i casi FTC degli ultimi anni contro Kochava (2022) e X-Mode (2024), data broker citati in giudizio per la vendita di dati sensibili dei consumatori come l'accesso a cliniche ginecologiche e di dati di localizzazione grezzi senza il consenso dell'utente. Anche qui, una citazione dal New York Times (2012), aiuta a chiarire la portata del fenomeno: "that's not my phone. It's my tracker".

La competizione tra le aziende è influenzata dalla quantità di informazioni sui consumatori che possono acquisire dai data broker.

Un aspetto cruciale riguarda l'impiego secondario dei dati dopo la loro raccolta iniziale. A questo riguardo, va evidenziato che i modelli di business dei data broker sono poco trasparenti e portano con sé implicazioni legali, etiche e sociali fondamentali.

Come spesso accade, le leggi tendono a restare indietro rispetto al progresso tecnologico. Pertanto, il settore dei servizi di brokeraggio dati è affetto da controversie, ciò può attribuirsi all'opacità delle attività dei data broker. Si auspica che, in tempi brevi, si assista a un rafforzamento o modifica delle norme nazionali, ma soprattutto internazionali, a riguardo.

SEZIONE 1- LETTERATURA

1.1 DEFINIZIONE DI DATA BROKER

La definizione di "data broker" (DB) è oggetto di contestazione. Non esiste una definizione univoca né in Europa né negli Stati Uniti.

Secondo l'ACCC (*Australian Competition and Consumer Commission*), un data broker è un fornitore che raccoglie informazioni personali e non, e le vende e/o le condivide con altri.

Ai fini di questo elaborato di tesi, si utilizzerà la seguente definizione di "data broker" che si allinea alle definizioni comuni negli Stati Uniti e in Europa.

DEFINIZIONE	SPIEGAZIONE
Un DB è un'azienda o unità commerciale	I data broker possono essere sussidiari di aziende più grandi.
che guadagna il suo reddito primario	Molte aziende guadagnano una parte dei ricavi fornendo dati, ma i data broker guadagnano la loro principale fonte di reddito fornendo dati.
fornendo, su gruppi di persone, inferenze o dati	Molti data broker vendono non solo dati, ma anche inferenze o previsioni su persone o gruppi di persone.
raccolti anche da fonti diverse dai soggetti stessi dei dati.	La quasi totalità dei DB non raccoglie i dati interamente e direttamente dai consumatori.

È importante notare che questa definizione, pur essendo ampia, esclude diverse grandi aziende molto influenti. Ad esempio, Amazon e Meta (Facebook, Instagram) interagiscono direttamente e quotidianamente con la maggior parte degli utenti di Internet in tutto il mondo. Ciò conferisce a queste aziende una conoscenza approfondita dei comportamenti e degli interessi dei propri utenti e possono consentire, ad esempio, ad inserzionisti di indirizzare gli annunci sulla loro piattaforma ai segmenti più inclini a rispondere positivamente. Tuttavia, non sono propriamente "data broker" secondo la definizione appena citata per due ragioni:

1. i loro maggiori introiti non sono imputabili alla vendita diretta di dati e,

2. i loro dati provengono quasi esclusivamente direttamente dai consumatori, piuttosto che essere forniti da altre aziende, da informazioni estratte da iscrizioni a periodici, da altri DB, documenti governativi, ...

Allo stesso modo, Disney non è un data broker, nonostante condivida i dati dei suoi clienti con una gamma di altre aziende, perché tali dati non sono la fonte primaria di ricavi dell'azienda. (Rieke, 2016)

1.2 LE FONTI DEI DATI DEI DB

I DB estraggono dati su ogni ente privato e pubblico, consumatori, aziende di ogni dimensione, governi, sia da fonti online che offline. Di seguito sono riportati le principali fonti **online** da cui i DB attingono per ottenere informazioni:

- Molti data broker raccolgono e organizzano i dati pubblici e facilmente accessibili, tipicamente utilizzando "web crawler" (programmi software progettati per raccogliere automaticamente dati da Internet) o li acquistano da altri data broker specializzati nella digitalizzazione di particolari tipi di record. Questi dati includono:
 - Documenti governativi, compresi rapporti sulla proprietà, documenti giudiziari, condanne penali e licenze professionali. Gli enti pubblici condividono una quantità significativa di dati.

Ad esempio, negli Stati Uniti, i dipartimenti statali della motorizzazione civile possono vendere dati a società private a fini di verifica dell'identità; alcuni stati consentono l'acquisto e la vendita dei registri elettorali. In Europa, la disponibilità di registri pubblici per ulteriori elaborazioni da parte dei data broker varia. Alcuni paesi scandinavi, ad esempio, sono notoriamente trasparenti rispetto ai registri fiscali delle persone in base alle leggi sulla libertà di informazione, mentre tali dati non sono disponibili al pubblico altrove. In Italia, per motivazioni analoghe, sono disponibili, ad esempio, dati catastali, dati sui procedimenti giudiziari, o dati sui cittadini elettori.

- Media, social network e dati online, comprese le informazioni pubbliche provenienti da LinkedIn, Facebook, Twitter e YouTube e siti di discussione con cui i DB collaborano frequentemente (sulle collaborazioni i DB fondano grande parte del loro business).
- Raccolgono dati relativi agli spostamenti fisici delle persone tramite strumenti come telecamere di sicurezza, sensori di movimento (IoT). A tal proposito, si approfondirà più avanti il "caso Kochava".
- Dati comportamentali: raccolgono anche dati di utilizzo online che si riferiscono a un dispositivo (un browser, una app, un ID mobile). In generale, i dati comportamentali digitali sono informazioni raccolte sui siti/app/eventi televisivi con cui un utente di un dispositivo interagisce nel tempo.

Da fonti **offline**, sono quasi interamente PII (personal identifiable information, informazioni personalmente identificabili), invece, nello spazio digitale, gran parte dei dati sono Non-PII, riferiti, appunto a un dispositivo piuttosto che a un individuo.

Le fonti di dati offline di interesse e disponibili per un data broker variano a seconda del tipo di data broker, spesso in base alle leggi del settore, così come, naturalmente, al tipo di dati che può essere più o meno prezioso per i DB stessi.

In generale, i dati offline provengono dalle seguenti fonti:

- Registri pubblici e informazioni pubblicamente disponibili: i registri pubblici sono una fonte comune di dati offline per gran parte dei data broker. Negli Stati Uniti, si ritiene che certe azioni governative, licenze e registrazioni dovrebbero essere disponibili per tutti da consultare, sia individui che organizzazioni. Questi includono procedimenti giudiziari, proprietà immobiliari, dati censuari e molto altro; oppure elenchi telefonici, elenchi professionali e listini, registri di veicoli, registri delle patenti di guida, registri di nascita e morte sono solo alcuni degli elenchi/registri da cui i DB estraggono innumerevoli informazioni.
- Per quanto riguarda il registro degli elettori, citato anche precedentemente, in alcuni stati, i record degli elettori possono essere utilizzati generalmente per scopi legati alle elezioni da parte di candidati e partiti, e se messi in collegamento con i dati di Meta (ex Facebook), possono generare episodi come lo scandalo di Cambridge Analytica.
- Questionari: le indagini sui consumatori sono condotte da società specializzate e da aziende che offrono servizi direttamente ai consumatori, ad esempio, si potrebbe pensare a dei produttori che includono un breve questionario al momento della registrazione del prodotto, spesso per fini legati alla garanzia. Le aziende che conducono queste indagini informano i partecipanti sul possibile utilizzo delle risposte: se rimarranno esclusivamente in loro possesso o se saranno condivise con terze parti come i data broker.
- Da aziende che lavorano in svariati settori, tra cui finanza, media, editoria, cataloghi, vendita al dettaglio, viaggi, intrattenimento, automobilistico. Tali aziende concedono in licenza alcune, o tutte, le informazioni dei loro clienti per vari scopi, inclusi rapporti sui consumatori, mitigazione del rischio e marketing.

I dati provenienti dalle tipologie di fonti offline appena citate possono essere di granularità diversa; ad esempio, l'iscrizione di un consumatore ad una certa rivista, oppure informazioni dettagliate sulla vendita di un'auto, servizi e riparazioni.

Infine, vi sono i dati modellati (o derivati): i data broker "creano" dati attraverso attività di modellizzazione statistica utilizzando alcune o tutte le suddette fonti di dati. I dati vengono modellati quando, ad esempio, gli elementi di dati desiderati non sono altrimenti disponibili, oppure per costruire previsioni. Questi processi analitici variano da semplici derivazioni a sofisticate analisi predittive. (Evan Selinger, 2018)

Data Collection Online & Offline

As consumers go about their business, data brokers may collect information about them.

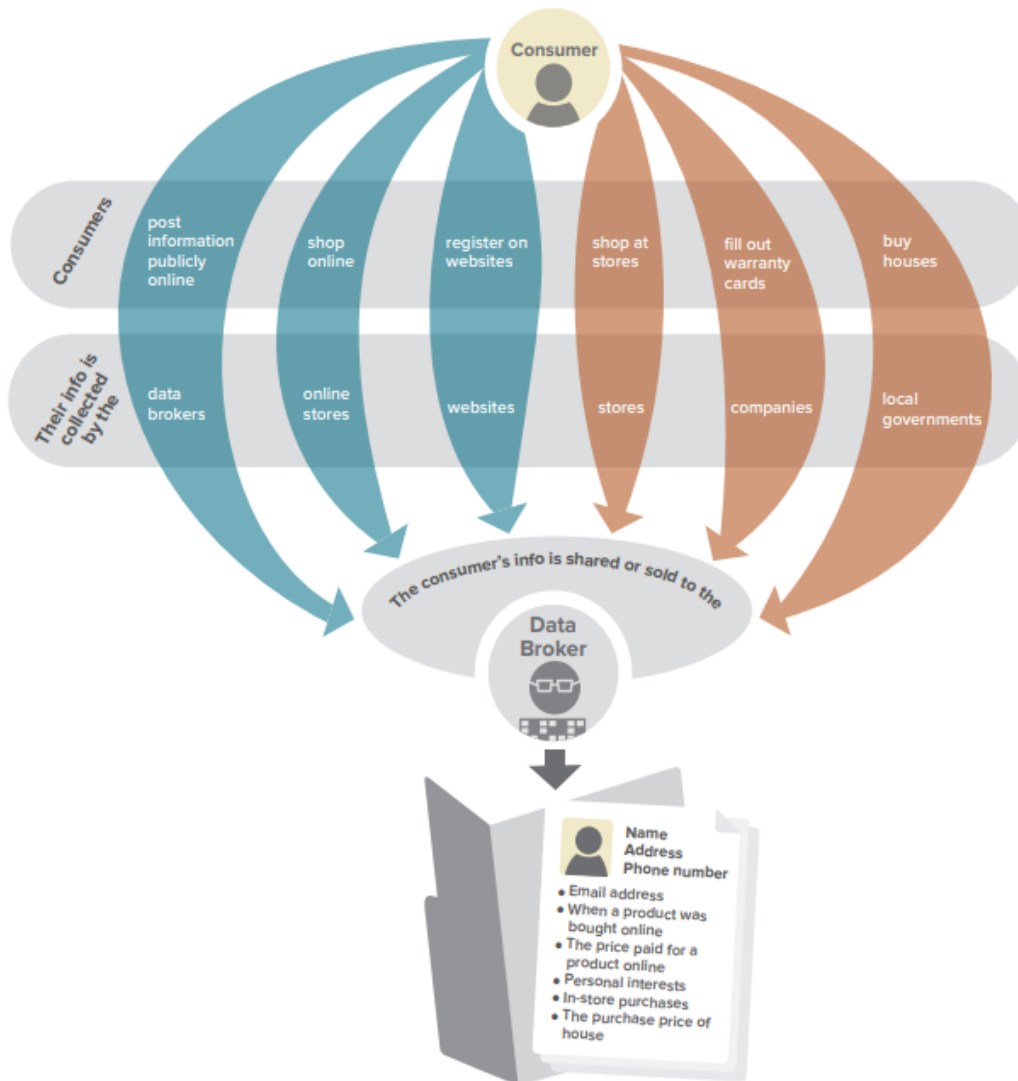


FIGURA 1: fonti per la profilazione (Federal Commission Trade (FTC), 2014)

1.3 IMPORTANZA DEI DB

Nel 2022, il mercato del brokeraggio di dati ha raggiunto un valore di 247 mila milioni di dollari. Si prevede che entro il 2028 questo mercato raggiungerà i 407 mila, mostrando un tasso (previsionale) di crescita annuale composto (CAGR) dell'8,67% (Maia Research, 2023).

Le suddette informazioni evidenziano una crescita significativa del mercato, crescita che evidenzia l'importanza dei servizi di brokeraggio di dati nell'economia digitale globale. L'aumento dell'adozione di tecnologie avanzate per l'analisi dei dati, la crescente

digitalizzazione delle imprese e l'incremento della domanda di soluzioni basate sui dati sono alcune delle cause principali di questa crescita.

In entrambe le definizioni citate nel primo paragrafo e, in generale, un aspetto comune delle svariate definizioni dei DB è che sono considerati dei "fornitori". Fa parte della conoscenza comune quanto ad oggi i dati siano importanti per il business di tutte le aziende. Sono così importanti da essere definiti con il valore di **forniture**. Qualsiasi tipo di azienda ha bisogno dei dati, dalle big-tech alla piccola impresa artigianale.

Nonostante i DB vendano enormi quantità di dati ad ogni tipologia di business, non godono della risonanza delle società sopracitate, eppure la loro rilevanza nella società odierna e futura è evidente, ed è necessario comprendere le dinamiche che regolano il loro mercato per permetterne il corretto funzionamento.

Nell'ultima decade le autorità antitrust hanno sviluppato varie inchieste sulla figura dei DB. Le più importanti sono:

- "Data Brokers: A Call for Transparency and Accountability" (2014) della Federal Commision Trade.
- "Update report into adtech and real time bidding" (2019) dell' Information Commissioner's Office.
- Infine, "Digital Platform Services Inquiry [...] report on data broker" condotta dall'ACCC, l'antitrust australiano, il cui obiettivo è:

"fornire una relazione [...] su diversi servizi di piattaforma digitale. I beni e servizi inclusi nell'Ordinanza del Tesoriere per l'indagine sono:

(a) servizi di piattaforma digitale

(b) servizi di pubblicità digitale forniti dai fornitori di servizi di piattaforma digitale

(c) servizi di raccolta, archiviazione, fornitura, elaborazione e analisi dati forniti da:

(i) fornitori di servizi di piattaforma digitale; o

(ii) data broker."

Questo fa comprendere il crescente interesse verso i DB delle autorità incaricate di garantire e regolare la concorrenza dei mercati.

1.4 TIPOLOGIA DI DB

È possibile classificare i DB secondo le tipologie di dati che raccolgono. Si identificano, dunque:

- **first-party data broker:** imprese che raccolgono dati sui propri consumatori (first-party data) che li analizzano e utilizzano internamente per migliorare la qualità del proprio servizio, discriminare, sviluppare campagne marketing e/o vendono o condividono con altri. Sono chiamati “broker di prime parti” perché hanno una relazione diretta con il cliente, anche se è difficile che essi rientrino nella definizione iniziale di DB; e
- **third-party data broker:** imprese che raccolgono dati sui consumatori da una varietà di fonti di terze parti e vendono o condividono tali dati con altri. (ACCC, 2023) e (Burdova, 2020).

I medesimi dati possono rientrare in entrambe le categorie “first” e “third-party data”; tuttavia, alcuni appartengono più frequentemente a una o all'altra.

I first-party data più comuni sono:

- Informazioni demografiche: età, istruzione, occupazione, stato civile
- Interessi: prodotti individuali, categorie di prodotti, materiali di marketing preferiti, contenuti preferiti
- Posizione: dove gli individui vivono o lavorano, o dove si trova un'azienda
- Cronologia degli acquisti: prodotti acquistati, pacchetti acquistati, abbonamenti, vendite incrociate o promozioni aggiuntive, valore degli acquisti, da quanto tempo qualcuno è cliente, ...
- Coupon: clienti che hanno o non hanno utilizzato coupon, clienti che utilizzano esclusivamente coupon, clienti più inclini a utilizzare coupon
- Social media: piattaforme su cui un cliente è più attivo e dove interagisce più spesso con il tuo marketing (Swendeman, 2023)

I third-party data più frequenti, invece, sono:

- Rapporti con i media (abbonamenti a riviste, appassionati di fumetti)
- Reddito familiare
- Hobby (ciclismo, caccia, maglia, ecc.)
- Occasioni (laurea, cambio di residenza)
- Abitudini di acquisto

(Langfelder, 2023)

Google è un data broker? Il gigante di Mountain View non può rientrare nella definizione di DB precedentemente fornita, in quanto la maggior parte del suo fatturato è dovuto al servizio di pubblicità (Casarella, 2023); tuttavia, vista l'enorme quantità di dati che possiede, può essere portato come esempio di un'azienda che opera sia come first sia come third-party data broker.

Come first-party DB:

- **Google Search e Google Account**, si intende la ricerca Google, Gmail, YouTube, Google Maps e altri. Google raccoglie una vasta quantità di dati direttamente da queste interazioni dei propri clienti. Questi dati includono le query di ricerca, le e-mail, i video guardati, le posizioni geografiche e altro ancora. Google utilizza questi dati per personalizzare l'esperienza dell'utente, fornire risultati di ricerca pertinenti, offrire annunci mirati, vendendo i dati agli inserzionisti.

Come third-party data broker:

- **Google Ads** permette agli inserzionisti di mostrare ai consumatori annunci su vari siti web, app e altri luoghi virtuali. Google utilizza i dati raccolti da molteplici fonti, sia interne che esterne, per consentire agli inserzionisti di segmentare specifici gruppi demografici in base a interessi e comportamenti degli utenti. Quindi raccoglie informazioni dagli utenti (anche) per i propri clienti.
- **DoubleClick** una piattaforma che offre strumenti per la gestione della pubblicità digitale e permette un tracciamento avanzato degli utenti, al fine di mostrare annunci pubblicitari sempre più personalizzati ed in linea con il "profilo" dell'utente. Offre servizi come la pubblicità display, cioè tramite siti web, blog, app mobili e altri spazi digitali sottoforma di banner, immagini, animazioni, ... o la pubblicità tramite video.

Infine, c'è un caso in cui Google offre un servizio da first and third party-data broker contemporaneamente:

- **Google Analytics**; questo servizio consente ai proprietari di siti web di raccogliere e analizzare i dati sull'utilizzo del proprio sito in maniera rapida e agevole. Quando i proprietari di siti implementano Google Analytics sul proprio sito, Google raccoglie dati sull'attività degli utenti per conto di quei siti (third-party). I dati raccolti appartengono al proprietario del sito, tuttavia, Google può chiederne l'accesso

(usando questi dati da first-party DB). Di norma, l'accesso viene consentito in quanto Google offre un servizio di maggior qualità a chi permette la condivisione.

Un'altra tipologia di DB sono i **personal data broker**:

I personal data broker (PDB) nel mercato dei dati personali offrono ai consumatori una remunerazione per i dati che hanno creato presso i fornitori di contenuti e servizi online (Content and Service Providers- CSP come Facebook, Instagram, ...). La generazione di queste informazioni può essere intesa come un servizio secondario del loro utilizzo dei CSP.

L'idea dei mercati dei dati personali risale a Laudon (1996), che immaginava la creazione di un mercato nazionale dell'informazione, in cui i soggetti che avessero generato dati potessero depositare le proprie informazioni presso istituzioni simili a banche e ricevere un compenso per l'uso dei loro dati. Nell'odierna economia dei dati, i personal data broker potrebbero essere considerati le "banche" previste in questa idea, poiché consentono ai consumatori di monetizzare i loro dati. Nella stragrande maggioranza dei casi, invece, il valore dei dati personali che i consumatori creano e rivelano come "servizio secondario" dell'uso dei servizi dei CSP è catturato da questi ultimi. Di conseguenza, società come Meta o Google basano i loro modelli di business principalmente sulla raccolta di dati degli utenti e sulla monetizzazione, più o meno diretta, di essi.

I PDB si impegnano nella raccolta, archiviazione e aggregazione dei dati provenienti da numerosi utenti, con l'obiettivo di raggiungere una posizione competitiva rispetto ai principali CSP per quanto riguarda i ricavi derivanti dai dati; tuttavia, la forza di mercato dei PDB non è equiparabile a quella dei DB e, in realtà, il loro modello di business non sembra essere competitivo nella catena di fornitura dei dati. (Krämer, et al., 2018)

1.5 DATI E SERVIZI SU CUI I DB FONDANO IL LORO BUSINESS

Nel mercato dei data broker, i dati su cui i DB fondano il proprio business sono:

1. Marketing

Raccolgono e/o vendono informazioni sui comportamenti di navigazione, acquisti passati e interessi. Sono responsabili del "marketing personalizzato" così come della maggior parte degli annunci online.

2. Informazioni finanziarie

a. Proprietà immobiliare

Raccolgono informazioni dalle banche dati catastali per ottenere dettagli completi sulle proprietà immobiliari e sulla solvibilità dei proprietari. Queste informazioni possono essere

utilizzate per valutare il valore delle proprietà e determinare l'affidabilità creditizia degli individui.

b. Reddito

Raccolgono dati sul reddito personale per fornire profili dettagliati ai potenziali clienti. Queste informazioni sono cruciali per valutare la capacità di un individuo di ottenere prestiti o mutui. Inoltre, aggregano varie informazioni finanziarie personali, sia storiche che attuali, per le società di credito o le banche, contribuendo alla costruzione del “credit score”, che determina, appunto, l'idoneità per l'ottenimento di un prestito.

3. Selezione del personale

Compilano documentazioni e utilizzano informazioni personali per offrire servizi di esame della formazione dei candidati al personale “risorse umane” degli svariati clienti.

4. Mitigazione del rischio

Raccolgono e/o vendono una varietà di informazioni sui precedenti penali, patrimoniali ecc., per fornire report di valutazione a varie società commerciali e di investimento.

5. Informazioni sulla salute

Raccolgono informazioni sulla salute generale e/o le vendono a società nel settore sanitario. Queste informazioni possono essere utilizzate per indirizzare gli utenti con annunci di prodotti per la salute e persino stabilire le tariffe assicurative. (2023)

Che servizi offrono i data broker?

- **“Original lists”**, ovvero, elenchi di consumatori raggruppati in base a caratteristiche o comportamenti previsti: sono elenchi di persone e informazioni di contatto basati su determinati criteri, comunemente utilizzati nel marketing. Ad esempio, un data broker potrebbe vendere un elenco di “uomini che vivono a New York e che hanno abbonamenti ai campi da golf”; oppure, un DB statunitense, Experian, ha venduto, per il mercato USA, un segmento chiamato “Thrifty Elders” (anziani parsimoniosi), che comprende “single tra i 60 e i 70 anni con reddito basso”. Nei Paesi Bassi ha offerto dati nella categoria “less successful” (i meno riusciti), ovvero “persone che stanno molto a casa” o che vivono in “case fatiscenti”. I data broker segmentano spesso i consumatori in base ai principali accadimenti della vita, come: il matrimonio, l'acquisto di una casa o l'inizio di una carriera universitaria di un figlio.

- **Punteggi:** Previsioni sul comportamento dei consumatori basate sui dati relativi a quella persona, utilizzate per un'ampia gamma di scopi. Ad esempio:
 - i punteggi di credito prevedono la probabilità che una persona non riesca a ripagare un prestito;
 - i punteggi di marketing possono prevedere la probabilità che una persona acquisti un determinato prodotto;
 - i punteggi lead potrebbero prevedere la probabilità che un potenziale consumatore sia un cliente redditizio.
 - I punteggi di frode potrebbero prevedere la probabilità che una particolare transazione sia fraudolenta. E i punteggi di stress potrebbero aiutare i clienti delle organizzazioni a gestire i costi e i rischi sanitari.
- **Rapporti sui consumatori:** Fascicoli su persone particolari e identificabili, comunemente utilizzati per valutazioni personalizzate su crediti, assicurazioni, impiego o simili. I rapporti sui consumatori spesso contengono dati reali, ma verranno utilizzati per fare deduzioni (per esempio, sotto forma di punteggi di credito).
- **“Look-alike models”:** Modelli che utilizzano dati noti su una persona per prevedere il suo comportamento in base al comportamento di persone in situazioni simili di cui il data broker dispone già di dati. Questi modelli sono comunemente usati nel marketing.
- **Implementazione dei dati:** Se i clienti dei DB dispongono di certi dati sui propri consumatori (es. e-mail), ma desiderano creare profili più completi, i DB forniscono informazioni aggiuntive su quella persona (ad esempio, sul recente modello di acquisto) (Rieke, 2016).

1.6 CHI SONO I PRINCIPALI DB

I principali players nel mercato della fornitura dei dati (**third-party data broker**) sono:

Acxiom



FIGURA 2: logo Acxiom (ACXIOM)

La Mission di Acxiom è “Le persone sono complesse e multiformi, quindi se vuoi connetterti con loro, devi comprendere l'intero quadro” (ACXIOM)

È uno dei principali DB e il suo focus è sul marketing. Acxiom aiuta le imprese a conoscere meglio i propri clienti grazie all'enorme quantità di dati che è riuscita ad ottenere con gli anni e che continua ad estrarre quotidianamente. Nel 2022, LiveRamp, la società che ha acquisito Acxiom nel 2014 e che offre servizi correlati al brokeraggio di dati, ha riportato ricavi di circa 588 milioni di dollari, riflettendo un aumento del 21% rispetto all'anno precedente.

Acxiom è stata fondata nel 1969 da Charles D. Ward e Clinton T. O'Brien con il nome di Demographics Inc., l'azienda ha iniziato come fornitore di informazioni demografiche per le elezioni politiche. Nel corso degli anni, Acxiom ha ampliato il proprio focus per includere una vasta gamma di servizi legati alla gestione dei dati, all'analisi e al marketing personalizzato diventando, di fatto, uno dei primissimi DB; in questo modo, ha acquisito un notevole vantaggio nei confronti dei propri clienti poiché disponeva di una conoscenza molto più approfondita dei clienti finali rispetto alle stesse aziende.

Negli anni '80, l'azienda ha iniziato a utilizzare la tecnologia informatica per raccogliere e analizzare dati demografici, comportamentali e di consumo, diventando uno dei primi pionieri nel settore della gestione dei dati. Nel corso del tempo, Acxiom ha continuato a crescere attraverso acquisizioni strategiche e sviluppo interno, ampliando la sua presenza globale e il portafoglio di servizi offerti. Ha guadagnato un ruolo di rilievo come leader nel settore, offrendo soluzioni all'avanguardia per consentire alle aziende di ottenere una comprensione più profonda dei propri clienti, di raggiungere il pubblico appropriato e di perfezionare le loro strategie di marketing. La società è coinvolta in una vasta gamma di settori, tra cui retail, finanza, sanità, viaggi e molto altro, offrendo soluzioni personalizzate per le esigenze specifiche di ogni cliente.

Acxiom offre principalmente un servizio di “martech” (MARKeting-TECHnology) ai propri clienti: “nell'economia digitale di oggi, i marchi dipendono dalla tecnologia di marketing per automatizzare i propri flussi di lavoro e fornire, su larga scala, le esperienze personalizzate del marchio apprezzate dai loro clienti. Ma non tutti i brand hanno le capacità e le competenze necessarie per implementare, gestire e ottimizzare il proprio portafoglio di soluzioni di marketing. Di conseguenza, molti marchi scelgono di affidarsi a fornitori di servizi martech esterni. Ad oggi il 65% dei brand con un fatturato annuo superiore a 100 milioni di dollari collabora con tre o più fornitori di servizi martech per l'implementazione, l'ottimizzazione e/o la gestione dei servizi”.

Acxiom si presenta come una delle più grandi aziende che vendono servizi di martech e collabora con le più importanti aziende in tutti i campi; infatti, riesce ad offrire un servizio di altissima qualità grazie soprattutto alle collaborazioni strategica che ha ottenuto negli

anni. “collaborare con i migliori significa crescita per i nostri clienti. Serviamo come consulenti strategici ed esperti tecnici sulle più recenti tecnologie di marketing e integriamo i nostri prodotti di dati e identità in soluzioni cloud.”

Si parla di collaborazioni con alcune delle più importanti società che hanno a che fare con dati personali (e non) dei propri clienti, ad esempio: Adobe, Ebay, Google Cloud, Google Marketing Platform, LinkedIn, Meta, Snapchat, Spotify, Twitter.

Il servizio più innovativo offerto da Acxiom è il “**Customer Intelligence Cloud**” (CIC), una suite connessa di servizi e una piattaforma cloud di marketing integrato; sono soluzioni per l’analisi di: identità, acquisizione, mantenimento e “crescita” dei clienti. Fondamentali per il CIC sono le “**clean rooms**”; questo concetto viene utilizzato nell’industria manifatturiera e nella comunità scientifica per riferirsi a un luogo privo di contaminazioni. Le comunità del marketing e della tecnologia hanno adottato il termine “data clean rooms” e ne hanno ampliato il significato per includere:

- una zona “sicura e affidabile” in cui aziende e partner possono condividere dati;
- un data hub per la condivisione degli identificatori in un’area protetta e controllata;
- un ambiente che pulisce e unifica i dati utilizzando una chiave di identità comune;
- una zona neutrale con severi controlli di accesso che alimentano co-marketing, monetizzazione, analisi e reporting.

Acxiom afferma che, se si mettono insieme le ultime notizie e normative recenti (trattate nella seconda sezione di questo elaborato), non c’è da meravigliarsi che le clean rooms dei dati stiano emergendo come argomento di tendenza. Hanno il potenziale per risolvere molteplici problemi per i professionisti del marketing che desiderano esplorare e implementare questo concetto.

I progressi nelle tecnologie di condivisione dei dati, che possono essere utilizzate nelle clean rooms, consentono inoltre ai proprietari dei dati di controllare quali dati possono essere condivisi con i partner e per quale scopo specifico.

In conclusione, in un mondo in cui le normative e le pratiche sui dati sono in continua evoluzione, le data clean rooms e gli spazi di collaborazione offrono opportunità agli esperti di marketing e ai marchi di trovare modi nuovi e innovativi per lavorare insieme.

L’obiettivo di Acxiom è dare supporto ai propri clienti nel:

- **Acquisizione di nuovi clienti**

Con le pressioni economiche e il cambiamento del comportamento dei consumatori, i marchi sono più concentrati che mai nell’aiutare efficacemente le persone a trovare i loro marchi. Raggiungere gli obiettivi di marketing di acquisizione per stimolare la crescita è

sempre più un obiettivo chiave per i marchi. L'acquisizione di nuovi clienti continua a essere una delle principali sfide.

- **Mantenimento dei clienti**

Non vi è alcuna garanzia che i clienti rimangano tali col passare del tempo. Il marketing intelligente per la fidelizzazione dei clienti inizia dalla comprensione dei clienti, oggi e domani. Ciò richiede una solida base di dati e un vero approccio CX basato sui dati, che riunisca l'intelligence del momento attraverso i canali e che tenga in conto anche i dati storici. Le relazioni si costruiscono nel tempo e quando il percorso del cliente dura anni e decenni, il marchio ha bisogno di sviluppare una grande memoria insieme a una visione del momento.

- **Crescita dei clienti:** "Le persone amano comprare, comprare di nuovo e comprare di più dai marchi di cui si fidano".

Quando si tratta di creare CX di alta qualità, i marchi devono considerare lo scambio di valore. Se i clienti condividono i propri soldi, il loro tempo e le loro informazioni, si aspettano che tutte queste informazioni siano utilizzate in modo saggio e a loro vantaggio, offrendo esperienze migliori, più personalizzate e attuali ovunque scelgano di interagire con il marchio. Si tratta di fiducia nel marchio. I clienti si fidano dei brand che dimostrano di comprendere le loro esigenze e di interagire con loro in modo rispettoso, inclusi messaggi, cadenza e canale preferito.

Experian



FIGURA 3: logo Experian (EXPERIAN, 2024)

Experian è una delle principali società di servizi di informazione creditizia e di gestione del rischio al mondo. Essa svolge un ruolo chiave nel settore finanziario globale. Fondata nel 1996 con sede a Dublino, Irlanda, Experian ha costruito nel corso degli anni una vasta rete internazionale di operazioni. Attualmente, l'azienda è presente in oltre 37 paesi e fornisce i suoi servizi a clienti in più di 80 mercati in tutto il mondo, attestando la sua portata globale e la sua influenza nel settore dei servizi finanziari e oltre. Experian nel 2023 ha registrato un fatturato di 6,619 Miliardi di dollari. La sua presenza su scala globale consente ad Experian di offrire una gamma diversificata di soluzioni e servizi adattati alle esigenze specifiche dei mercati locali, mantenendo al contempo standard elevati di qualità e affidabilità in tutti i suoi servizi.

Basa il suo business sui dati di credit score sia di persone fisiche che di imprese.

Per quanto riguarda gli individui, l'obiettivo dichiarato di Experian è "aiutare le persone a ottenere una visione più chiara della propria situazione finanziaria, ad assumerne e a mantenere nel tempo una salute finanziaria ottimale".

Per le imprese, l'obiettivo è "fornire un servizio di consulenza per prendere decisioni più rapide e intelligenti, nell'erogare prestiti in modo responsabile, equo e tempestivo, riducendo al contempo il rischio e migliorando l'esperienza del cliente".

La mission di Experian: "[...] Il nostro obiettivo [...] è centrale per il nostro marchio, articolato dalle nostre persone e mutuamente rafforzato dalla nostra cultura. È la base di tutto ciò che facciamo, perché milioni di persone in tutto il mondo sono ancora escluse dall'accesso al credito equo e accessibile poiché sono invisibili al sistema finanziario. Vogliamo cambiare questo stato di cose. Quindi, è la nostra missione migliorare l'inclusione finanziaria, perché l'accesso al credito apre opportunità alle persone per trasformare le proprie vite." (2024)

CoreLogic



FIGURA 4: logo CoreLogic (CORELOGIC, 2024)

CoreLogic è un DB che si occupa di informazioni e analisi immobiliari e finanziarie, riconosciuto a livello globale per la sua leadership nel settore. La società è stata fondata nel 1991 con sede a Irvine, in California, CoreLogic ha costruito nel corso degli anni una solida reputazione per l'elevata qualità dei servizi di analisi di dati e del settore immobiliare e finanziario. CoreLogic USA ha registrato un fatturato di 1.64 Miliardi di dollari nel 2020, considerando anche le sedi in Australia, Nuova Zelanda e Regno Unito ha raggiunto circa i 2 Miliardi (Moody's, 2024).

CoreLogic opera a livello internazionale, offrendo un'ampia gamma di soluzioni innovative a clienti in tutto il mondo, con una presenza in oltre 40 paesi. Grazie alla sua vasta rete di dati e alle sue avanzate capacità analitiche, CoreLogic fornisce informazioni dettagliate e approfondite che supportano i clienti nel prendere decisioni informate e nella gestione dei rischi. L'azienda è ampiamente riconosciuta per i suoi servizi di valutazione immobiliare, gestione dei dati, analisi del rischio e prevenzione delle frodi. La sua presenza globale e la reputazione di affidabilità e precisione dei dati permettono la collaborazione con banche,

istituti finanziari, assicuratori, agenzie governative e altri attori del settore. CoreLogic “continua a guidare l'innovazione nel settore, investendo in tecnologie avanzate e sviluppando nuove soluzioni per affrontare le sfide emergenti nel mercato immobiliare e finanziario”.

“La nostra missione è rendere l'industria immobiliare più veloce, più intelligente e più orientata alle persone”

CoreLogic si occupa dell'industria immobiliare offrendo soluzioni su:

- **Real Estate**

- **Valutazione immobiliare**

CoreLogic fornisce servizi avanzati di valutazione immobiliare che aiutano agenti immobiliari, venditori, acquirenti e istituti finanziari a determinare il valore di una proprietà. Utilizzando dati accurati e sofisticati modelli analitici, CoreLogic offre stime precise che riflettono le condizioni di mercato locali, i trend di vendita e altre variabili influenti.

- **Analisi del mercato immobiliare**

La piattaforma di analisi di CoreLogic fornisce insights dettagliati sul mercato immobiliare, inclusi dati sulle vendite, tendenze dei prezzi, domanda e offerta di proprietà in una determinata area. Queste informazioni aiutano gli investitori, gli sviluppatori e gli agenti immobiliari a prendere decisioni strategiche e informate.

- **Soluzioni per l'elaborazione di transazioni immobiliari**

CoreLogic offre soluzioni software per semplificare e automatizzare i processi di transazione immobiliare, inclusi strumenti per la gestione di contratti, chiusure, titoli e altro ancora. Questi strumenti migliorano l'efficienza operativa e riducono i rischi associati alle transazioni immobiliari.

- **Mutui Ipotecari**

- **Valutazione del Rischio Ipotecario**

CoreLogic fornisce servizi di valutazione del rischio per aiutare le istituzioni finanziarie a valutare la solvibilità dei mutuatari e il rischio associato ai prestiti ipotecari. Utilizzando dati sul credito, dati immobiliari e modelli analitici avanzati, CoreLogic identifica i potenziali rischi e aiuta a mitigarli.

- **Soluzioni per la gestione del portafoglio di prestiti**

CoreLogic offre soluzioni software per la gestione dei portafogli di prestiti ipotecari, inclusi strumenti per la gestione dei rischi, la valutazione della qualità del credito e la conformità normativa. Queste soluzioni aiutano le istituzioni finanziarie a gestire in modo efficiente e sicuro i loro portafogli di prestiti ipotecari.

➤ **Servizi di elaborazione di mutui**

CoreLogic fornisce servizi di elaborazione di mutui che semplificano e accelerano il processo di sottoscrizione di mutui ipotecari. Questi servizi includono valutazioni immobiliari, verifica dei documenti finanziari, analisi del rischio e altro ancora.

▪ **Assicurazioni**

➤ **Valutazione dei rischi assicurativi**

CoreLogic offre servizi di valutazione dei rischi per le compagnie di assicurazione, utilizzando dati sulle proprietà, dati meteorologici, dati sulle calamità naturali e altri dati pertinenti. Questi servizi aiutano le compagnie di assicurazione a valutare e gestire i rischi associati alle polizze di assicurazione immobiliare.

➤ **Gestione delle richieste di risarcimento**

CoreLogic fornisce soluzioni software per la gestione delle richieste di risarcimento, che consentono alle agenzie di assicurazione di gestire in modo efficiente e accurato le richieste di risarcimento dai loro assicurati. Queste soluzioni aiutano a migliorare l'esperienza del cliente e a ridurre i costi operativi.

➤ **Analisi delle tendenze e previsioni per l'industria assicurativa**

Utilizzando dati e analisi avanzati, CoreLogic offre insights sulle tendenze del settore assicurativo e previsioni di mercato per aiutare le compagnie di assicurazione a prendere decisioni strategiche e a sviluppare prodotti assicurativi innovativi. (CORELOGIC, 2024)

Oracle



FIGURA 5: logo Oracle (ORACLE, 2024)

Per la configurazione che ha assunto Oracle negli anni, la definizione di DB utilizzata per questo elaborato è imprecisa. Inoltre, la società gode di una elevata notorietà, cosa che, di norma, non accade per i DB; tuttavia, la società nasce come DB e, tutt'oggi, continua a basare gran parte del proprio business sui dati; è tra le più grandi società mondiali, al punto da offrire servizi in moltissimi settori e in 4 continenti, Oracle ha registrato un fatturato nell'anno fiscale 2023 di 50 Miliardi (Oracle, fatturato record di 50 miliardi di dollari nell'anno fiscale 2023, 2023).

Nel 1977, con il nome di **Software Development Laboratories**, nasce come startup nel cuore della Silicon Valley, fondata dagli ingegneri Larry Ellison, Bob Miner e Ed Oates. Nel 1979, la società cambia nome in Relational Software Inc. per poi adottare definitivamente il nome **Oracle Corporation** nel 1982. Nel 1986, Oracle diventa una società quotata in borsa al

NASDAQ, con la sigla ORCL. Nel 1989, viene inserita nell'indice S&P 500 e, continuando a crescere molto velocemente, si trasferisce nella sua attuale sede a Redwood Shores, California, a causa delle dimensioni ormai troppo ridotte degli uffici precedenti. Nel 1992, Oracle festeggia i suoi 15 anni di attività e il rilascio di Oracle7, un database all'avanguardia che segna un importante traguardo per l'azienda. Nel 1995, il CEO Larry Ellison introduce una strategia innovativa per distribuire il software Oracle, sfruttando il potere emergente di Internet. Nel 1999, Oracle Database viene riconosciuto da InformationWeek come uno dei prodotti più influenti degli anni '90, definendo uno standard nel settore. Nel 2000, crea "Oracle E-Business Suite 11i", la prima suite di app aziendali totalmente integrata. Nel 2005, avvia un trend di acquisizione high-tech acquisendo PeopleSoft, un leader nelle app HR ed ERP. Nel 2008, Marvel Entertainment inizia a utilizzare le applicazioni Oracle per abbattere i silos e sfruttare al meglio i dati. Nel 2010, Oracle acquisisce Sun Microsystems, diventando l'amministratore di Java e consolidando la sua strategia per progettare soluzioni hardware e software integrate. Nel 2018, lancia il primo database autonomo del settore, con funzionalità di ottimizzazione, gestione e patch. Nel 2019, a Barcellona, ELEM utilizza il cloud altamente performante di Oracle per eseguire calcoli su modelli integrali di cuori malati, combattendo le malattie cardiache. Nel 2021, Oracle nell'ambito del COVID-19, supporta agenzie governative e comunità mediche nel comprendere e combattere il virus. Nel 2022, Oracle acquisisce il pioniere degli EHR Cerner, come parte di una missione più ampia per creare e implementare sistemi informativi sanitari integrati e automatizzati. Inoltre, Oracle Red Bull Racing difende il suo titolo di vincitore del campionato mondiale piloti di F1 e vince il campionato mondiale costruttori, affidandosi a Oracle Cloud per supportare la propria strategia di gara e coinvolgimento dei fan. (ORACLE, 2024). Infine, riesce ad offrire un elevatissimo numero di servizi a persone e imprese nei seguenti settori:

- **Settore automobilistico**

Oracle fornisce soluzioni per la gestione della supply chain che aiutano le aziende automobilistiche a ottimizzare la produzione, la distribuzione e la logistica dei componenti e dei veicoli stessi; fornisce anche soluzioni CRM (Customer Relationship Management) che consentono alle aziende automobilistiche di gestire le relazioni con i clienti in modo efficace; e offre soluzioni di analisi dei dati che consentono alle aziende di raccogliere, analizzare e utilizzare dati provenienti da diverse fonti, inclusi veicoli connessi, per migliorare l'efficienza operativa, prevedere la domanda del mercato e sviluppare prodotti più mirati.

- **Comunicazioni**
- **Ingegneria edilizia**

Tramite Oracle cloud, permette di connettere tutti i team di progetto, i processi e i dati per mantenere tutto sincronizzato.

- **Beni di largo consumo**
- **Formazione**

Oracle offre servizi ai diversi livelli di istruzione, ad esempio:

➤ **Istruzione superiore**

Analisi dei dati educativi: Oracle offre soluzioni per l'analisi dei dati educativi, che consentono alle istituzioni di analizzare e utilizzare dati sui risultati degli studenti, l'efficacia dei programmi didattici, il coinvolgimento degli studenti e altro ancora per migliorare il metodo di insegnamento.

➤ **Istituti di ricerca**

Accelerazione della ricerca utilizzando la più recente infrastruttura di cloud computing ad alte prestazioni.

➤ **Per entrambi**

Gestione accademica e amministrativa: Oracle fornisce soluzioni per la gestione accademica e amministrativa, che includono sistemi di gestione degli studenti, registrazioni, pianificazione delle lezioni, ...

Tecnologie per l'apprendimento online: offre soluzioni tecnologiche per l'apprendimento online, che includono piattaforme di e-learning, sistemi di gestione dei contenuti, strumenti di collaborazione online, e altro ancora per supportare l'apprendimento a distanza e ibrido.

- **Energia e acqua**
- **Servizi finanziari**

Oracle semplifica la ricerca automatica di prodotti e servizi finanziari per i clienti in base alle loro esigenze, preferenze e comportamenti; ovvero, in base al segmento di cui i clienti fanno parte.

- **Alimentare**
- **Governi**

Oracle offre un servizio di cloud che va dal garantire il comando e il controllo a livello tattico, a trasformare delle immagini satellitari in intelligence utilizzabile; Oracle Cloud è progettato per alti livelli di sicurezza con servizi autonomi avanzati, intelligenza artificiale integrata e flessibilità multi-cloud. Inoltre, offre piattaforme di gestione dei dati, sicurezza, integrazione, analisi e sviluppo per gestire missioni impegnative di sicurezza nazionale in tutto il mondo

- **Salute**

La soluzione cloud di Oracle per l'Electronic Health Record (EHR) agevola i team di assistenza sanitaria nell'accesso ai dati critici dei pazienti, migliorando i flussi di lavoro e agevolando il trasferimento di informazioni sui pazienti. Questa soluzione offre ai professionisti un accesso completo alla documentazione di ogni paziente, consentendo loro

di coordinare l'assistenza tra sistemi sanitari, prendere decisioni più informate e aumentare l'efficienza complessiva.

- **Alta tecnologia**
- **Servizi alberghieri**

Hotel e resort, Casinò e gaming, Food and Beverage nel settore alberghiero e Settore crocieristico

La tecnologia Oracle Hospitality connette vendite di eventi, camere degli ospiti e gestione, offrendo le integrazioni necessarie per migliorare la qualità offerta agli ospiti.

Permette la dotazione degli hotel di sistemi di property management e point-of-sale (POS) che offrono al personale migliori prestazioni. I property-management system (PMS) e le soluzioni POS per hotel di Oracle integrano facilmente altre tecnologie per i clienti, ad esempio, accelerare il check-in e il check-out, gestire le operazioni di house-keeping e manutenzione e utilizzare i dispositivi mobili per migliorare il servizio offerto ai clienti ovunque e in qualsiasi momento.

Implementa un sistema di property management per i casinò che riconosce e premia i grandi giocatori.

- **Produzione industriale**
- **Scienze della vita**
- **Media e intrattenimento**
- **Settore petrolifero e del gas**
- **Servizi professionali**
- **Pubblica sicurezza**

Oracle offre il servizio di gestione delle tecnologie più recenti, come body cam (sempre più sofisticate) e altri dispositivi di acquisizione dati, Internet of things (IoT) e analisi più approfondite. Tale servizio aiuta a garantire sia la sicurezza degli agenti che quella pubblica e consente alle forze dell'ordine di servire le comunità in modo più efficace.

Inoltre, permette alle forze dell'ordine di adottare un approccio preventivo anziché reattivo per garantire la sicurezza del pubblico: "per stare al passo con le tendenze della criminalità, le agenzie di pubblica sicurezza devono anticipare e rispondere a tali minacce".

- **Oracle Public Safety Dispatch Command Center** è una piattaforma CAD basata su cloud per la gestione e l'instradamento delle chiamate di emergenza.
- **Oracle Public Safety Wearable Computer System** offre comunicazioni indossabili agli agenti e sistemi di telecamere che si attivano automaticamente all'invio, registrando le interazioni tra le forze dell'ordine e i membri della comunità per favorire trasparenza, fiducia e responsabilità.

- **Oracle Public Safety Personal Communication System** è un'applicazione mobile su dispositivo sicuro che agevola l'invio e le interazioni sul campo tra i soccorritori.
 - **Oracle Public Safety Jail Management System** è una soluzione cloud per la gestione delle strutture carcerarie, progettata per eliminare doppioni di dati e migliorare l'efficienza e la sicurezza dei processi dalla prenotazione all'eventuale trasferimento o rilascio.
- **Retail**
 - **Viaggi e trasporti**
 - **Distribuzione all'ingrosso**

Illion



FIGURA 6: logo Illion (ILLION, 2024)

Illion, attualmente, non è un DB rilevante quanto i precedenti; tuttavia, ricopre una fetta importante nel mercato australiano. È il principale fornitore indipendente di prodotti e servizi di dati e analisi affidabili in Australia e Nuova Zelanda. Illion sfrutta i suoi registri di credito per consumatori e commerciali per fornire soluzioni complete di gestione ai clienti nei settori dei servizi finanziari, delle telecomunicazioni e delle utilities.

La storia di Illion, precedentemente conosciuta come Dun & Bradstreet, ha le sue radici nel 1841, quando l'azienda è stata formata negli States, si occupava di raccogliere, elaborare e distribuire dati commerciali su aziende e imprenditori. Nel corso degli anni, l'azienda ha continuato a crescere e ad espandersi, diventando un leader globale nel settore delle informazioni commerciali.

In Australia, l'azienda ha iniziato la sua attività nel 1986, sotto il nome di "D&B Australia", fornendo servizi di informazioni commerciali per aiutare le aziende a valutare il credito e gestire i rischi finanziari. Nel corso del tempo, l'azienda ha continuato ad evolversi e ad adattarsi alle mutevoli esigenze del mercato. Nel 2004 è stato lanciato il "consumer risk score" Nel 2005, Dun & Bradstreet Australia è diventata parte di Pacific Dunlop Limited. Nel 2018, la società ha cambiato il suo nome in "Illion", riflettendo il suo impegno a fornire soluzioni innovative e all'avanguardia nel campo delle informazioni commerciali e creditizie in Australia e Nuova Zelanda. Illion si è impegnata a "offrire dati e analisi affidabili per aiutare le imprese a prendere decisioni informate e gestire i rischi in modo efficace". Continua a essere un leader nel settore delle informazioni commerciali e creditizie,

offrendo una vasta gamma di prodotti e servizi che aiutano le aziende a comprendere meglio i propri clienti, gestire i rischi finanziari e prendere decisioni commerciali informate.

Essendo uno dei principali uffici creditizi per consumatori e aziende dell'area dell'Australia, i dati unici, il software, le analisi e le soluzioni di gestione del rischio costituiscono una parte integrante dell'ecosistema finanziario della regione. Conosciuta per l'esperienza nella gestione del rischio creditizio, Illion archivia e analizza i dati unici di milioni di individui ed entità commerciali. Fornisce anche una gamma completa di prodotti che coprono l'intero ciclo di vita del credito, tra cui la generazione di lead, il software decisionale, le tecnologie di pagamento e la gestione delle offerte.

L'appartenenza di Illion alla Rete Mondiale di Dun & Bradstreet, un'alleanza di fornitori commerciali di informazioni in 220 paesi, consente di combinare la conoscenza locale con dati globali su più di 500 milioni di aziende in tutto il mondo. L'obiettivo di Illion è quello di fornire ai clienti strumenti per prendere decisioni intelligenti e dati attentamente analizzati e derivati dal loro team.

Su cosa basa il proprio vantaggio Illion?

Esperienza comprovata - gestione di programmi di integrazione specializzati per le agenzie di informazioni garantendo la consapevolezza dei rischi, vincoli e sfide, nonché delle capacità per mitigarli e risolverli.

Copertura maggiore di quella richiesta: profondità, disponibilità e uptime che spesso supera le esigenze e i livelli di servizio richiesti dai clienti, garantendo operazioni senza intoppi per gli utenti.

Agenzia per i consumatori: l'agenzia per i consumatori di Illion si concentra sulla fornitura di dati rilevanti per i propri clienti, in particolare dati differenziali sui consumatori a rischio più elevato.

Accesso globale: accesso a dati aggiornati 5 milioni di volte al giorno su oltre 400 milioni di record aziendali globali provenienti da 220 paesi attraverso il piano di adesione alla Dun & Bradstreet Worldwide Network.

Leadership di pensiero nel settore: investimenti continui nell'innovazione dei prodotti e capacità di essere un consulente strategico su approcci tecnologici che sostengono la rivoluzione basata sui dati nei servizi finanziari.

Illion offre i seguenti servizi ai suoi clienti:

- **Controlla un'attività**

Sfruttando approfondimenti basati sui dati delle entità commerciali, Illion può aiutare a ridurre il rischio e consentire di prendere decisioni di credito più informate.

La mole di dati e le sofisticate analisi informano la gamma di report e punteggi di rischio predittivo proprietari. Questi report e punteggi rivelano informazioni utili sull'esposizione al rischio e sulle opportunità di crescita della azienda del cliente. Fornisce, inoltre, l'accesso a informazioni di terze parti come titoli fondiari e il registro dei titoli delle proprietà personali (PPSR), consentendo di convalidare o registrare la sicurezza per proteggere l'esposizione aziendale.

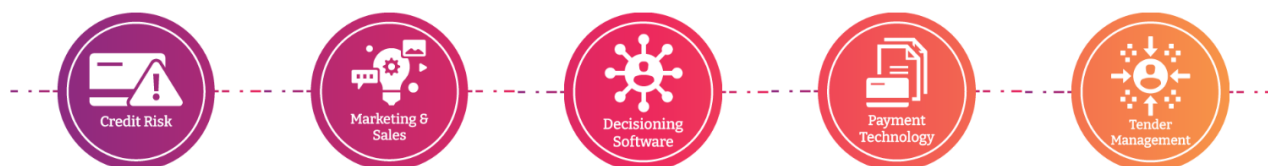


FIGURA 7: processo di analisi di Illion dei propri clienti

- **Rischio di credito**

Soluzioni di dati e analisi per proteggere l'azienda dai rischi

Dal 1986, Illion fornisce dati e approfondimenti per contribuire a rendere il credito accessibile a privati e imprese. Le soluzioni basate sulla tecnologia includono identificazione, reporting sul credito al consumo e commerciale, valutazione dell'accessibilità e della responsabilità e servizi di catena di fornitura. Queste soluzioni possono aiutare il cliente a prendere decisioni informate e tempestive in grado di ridurre al minimo i rischi, mitigare le perdite finanziarie e gestire le inadempienze.

- **Marketing e vendite**

Sbloccare potenti strategie di marketing e vendita con una gamma unica di prodotti di dati e analisi

Che si tratti di una conoscenza approfondita del comportamento dei clienti o di informazioni preziose sulla concorrenza, gli approfondimenti commerciali e sui consumatori "possono aiutare il cliente a scoprire nuove opportunità per portare l'attività al livello successivo".

- **Software decisionale**

Soluzione software decisionale che consente di personalizzare il processo di credito in base alle esigenze dell'azienda e dei clienti.

Illion Decisioning Service (IDS) è una soluzione di onboarding completamente automatizzata che consente la comunicazione diretta con i clienti nei punti critici durante tutto il processo di approvazione del credito, sia che si tratti di prestiti finanziari o automobilistici o di altri tipi di credito al consumo e commerciale.

- **Tecnologie di pagamento**

Tecnologie innovative per semplificare la riscossione dei pagamenti.

Per oltre 30 anni, Illion Digital Tech Solutions ha lavorato per rivoluzionare il settore dei pagamenti e degli incassi, servendo organizzazioni di servizi finanziari, governo, telecomunicazioni e servizi di pubblica utilità in tutto il mondo.

- **Gestione gare d'appalto**

Collegamento di acquirenti e fornitori in tutta l'Australasia

Illion TenderLink collega oltre 600 partner acquirenti con oltre 124.000 fornitori in tutta l'Australia e la Nuova Zelanda, fornendo assistenza ad acquirenti e fornitori nei settori governativo, pubblico e privato. Fondati nel 1994, da allora Illion è cresciuta fino a diventare uno dei più grandi mercati di gare d'appalto dell'Australasia. (ILLION, 2024)

Altri DB importanti sono: LiveRamp, Equifax, Nielsen, PropTrack, Quantum, Q.Checkout, TransUnion che offrono pressoché gli stessi servizi già esaminati per i DB precedenti.

Per quanto riguarda i PDB, invece, i principali player che ricoprono, o hanno ricoperto, una quota di mercato, sono:

Universal Basic Data Income (UBDI)

UBDI è una piattaforma di ricerca basata su un'app per smartphone. I partecipanti accettano di condividere alcuni dei loro dati personali, partecipare a sondaggi e vengono ricompensati per il loro feedback. Utilizzando il framework di digi.me per la condivisione dei dati, l'app ha integrazioni con centinaia di fonti di dati personali nei settori social, medico, finanziario, salute e fitness, musica e intrattenimento.

Datum

Datum è un PDB innovativo che offre una piattaforma basata su blockchain per la memorizzazione e la monetizzazione dei dati personali. Gli utenti possono controllare i propri dati, decidere quanto condividere e con chi, e ricevere una compensazione per la condivisione. La piattaforma utilizza il "token DAT" per gestire le transazioni e premiare gli utenti. Gli sviluppatori possono creare applicazioni che rispettano la privacy e utilizzano la blockchain di Datum per la sicurezza dei dati.

Datacoup

Datacoup è una piattaforma che consente agli utenti di controllare, gestire e monetizzare i propri dati personali. Gli utenti possono condividere i loro dati in modo sicuro e ricevere compensi in denaro, sconti o criptovalute, come Datum. Datacoup offre ai suoi client

- **Controllo dei Dati:** gli utenti possono scegliere quali dati condividere e con chi.

- **Monetizzazione:** gli utenti ricevono compensi monetari per la condivisione dei loro dati.
- **Sicurezza:** utilizzo di strumenti e app basati su blockchain per garantire la sicurezza dei dati.

1.7 MODELLI SUL MERCATO DEI DB: analisi della letteratura economica

In letteratura, il funzionamento del mercato dei DB è stato, ed è tuttora, analizzato per comprendere se vi sono fallimenti o malfunzionamenti che, in generale, contribuiscono alla mancanza di una corretta competizione.

Un primo lavoro (David Bounie, et al., 2021) indaga le strategie di un DB monopolista delle informazioni sui consumatori, che vende dati a due imprese simmetriche concorrenti tramite asta con esternalità negative: se un'azienda non acquista i dati dal DB, saranno acquistati dall'altra.

Avere più informazioni consente alle aziende di identificare meglio i consumatori e di discriminare il prezzo. Possono, dunque, estrarre più surplus, aumentando i loro profitti. D'altro canto, possedere più informazioni significa anche che la competizione tra le aziende sui consumatori che hanno identificato come appartenenti ai loro segmenti commerciali sarà, più forte. Questa maggiore competizione abbassa i profitti delle aziende. Nel complesso, il trade-off economico tra l'estrazione di surplus e l'aumento della competizione analizzato da questo articolo è quello in cui un DB vende strategicamente informazioni alle imprese con il fine di massimizzare il suo profitto.

Il modello prende in considerazione un data broker monopolista che vende i dati, o partizioni di essi a due imprese (indicate come 1 e 2), che vendono prodotti concorrenti ai consumatori. I consumatori sono uniformemente distribuiti su una linea unitaria $[0,1]$ come nel modello base di Hotelling. Il DB raccoglie informazioni sui consumatori, a un costo normalizzato a zero, che acquistano prodotti dalle imprese concorrenti. Le imprese possono acquistare informazioni dal data broker per fare discriminazione di prezzo tra i consumatori e vendono i propri prodotti, entrambe con costi di produzione normalizzati a 0. Le due imprese sono situate a 0 e 1 sulla linea unitaria.

Un consumatore situato in " x " ottiene un'utilità lorda " V " dal consumo del prodotto e affronta un costo di trasporto lineare con valore $t > 0$. Un consumatore acquista al massimo un'unità del prodotto e si assume che il mercato sia completamente coperto, cioè tutti i consumatori razionali acquistano il prodotto; p_1 e p_2 sono i prezzi fissati da 1 e da 2, rispettivamente. Un consumatore situato in x riceve la seguente utilità:

$$U(x) = V - tx - p_1, \text{ se acquista da 1,}$$

$$U(x) = V - t(1 - x) - p_2, \text{ se acquista da 2,}$$

$$U(x) = 0, \text{ se non acquista.}$$

Le imprese sanno che i consumatori sono distribuiti uniformemente sulla linea unitaria, ma senza ulteriori informazioni prima di acquistare informazioni dal DB. Non sono in grado di identificare le loro posizioni, dunque non sanno in che misura i consumatori apprezzino i loro prodotti (quanto siano “distanti” da loro) e, su di essi, non ha significato discriminare i prezzi.

Le imprese possono acquistare una “struttura di informazioni” da un data broker monopolista a un costo “ w ”. La struttura di informazioni consiste in una suddivisione della linea unitaria in n segmenti di dimensioni arbitrarie. Questi segmenti sono costruiti mediante unioni di segmenti elementari di dimensione $1/k$, come mostrato in figura 8, dove k è un intero esogeno che può essere interpretato come la qualità o la precisione delle informazioni. Infatti, minore è k , più i segmenti sono grandi e più difficile è riconoscere la posizione precisa dei consumatori; più grande è k , più precisi sono i dati venduti dal DB.

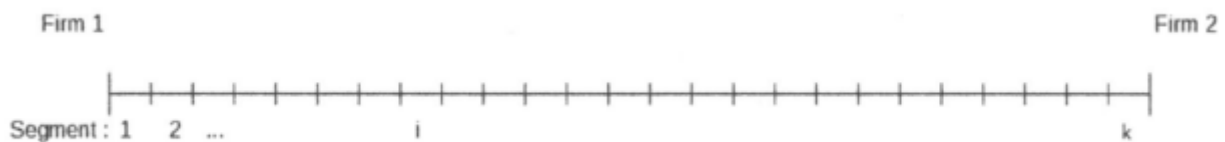


FIGURA 8: PARTIZIONE DI RIFERIMENTO

Il modello, inoltre, prevede la possibilità per il DB di vendere le informazioni su segmenti precisi della partizione, come mostrato in figura 9.

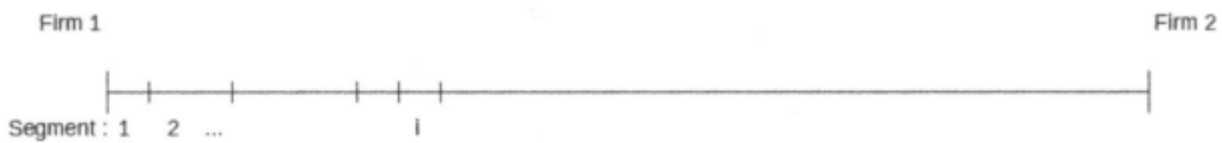


FIGURA 9: POSSIBILE SEGMENTAZIONE DEI DATI

Il DB può vendere anche segmenti combinati, cioè non “consecutivi” e questo permette di fare alcune considerazioni aggiuntive che “non sovrastimino l’effetto della competizione”; ciò sarebbe accaduto se il DB avesse venduto stesse partizioni alle imprese; infatti, l’articolo sostiene che consentire alle aziende di acquistare tutte le informazioni disponibili o nessuna informazione farebbe in modo che tra le imprese vi sia una forte competizione; si consideri, ad esempio, la situazione in cui un data broker venda alle imprese tutte le informazioni sui consumatori su una linea di Hotelling: le aziende competono per ogni consumatore su $[0,1]$, e la competizione è feroce, à la Thisse e Vives (1988), ma ciò non è detto sia la condizione che massimizzi il profitto del DB.

Si dimostra che il DB può aumentare i profitti vendendo la seguente partizione: le informazioni sui consumatori su $[0, x]$ all'impresa 1 situata in 0, e le informazioni sui consumatori su $[y, 1]$ all'impresa 2 situata in 1. Le aziende competono solo per i consumatori situati su $[x, y]$ e il DB sceglie ottimamente x e y al fine di ridurre la pressione competitiva delle informazioni, mentre massimizza l'estrazione di surplus dai consumatori.

Quando si parla di "ricombinare i segmenti", si intende la capacità di suddividere i consumatori in gruppi basati su caratteristiche comuni, in questo caso vicini, e poi ricombinare tali gruppi in modi diversi. Se i dati sui consumatori fossero utilizzati direttamente senza alcuna forma di segmentazione, potrebbero fornire un vantaggio competitivo a chiunque li possieda. Graficamente, i segmenti combinati sono rappresentabili come intervalli disgiunti, che possono portare a due effetti opposti sui profitti del DB.

- Da un lato, se gli intervalli disgiunti consentono alle aziende di colludere sul prezzo, imposteranno prezzi elevati per i consumatori, aumentando così la loro propensione a pagare per le informazioni.
- D'altro canto, gli intervalli disgiunti possono aumentare la pressione competitiva tra le aziende dando incentivi a catturare clienti "dall'altro lato" della linea di Hotelling, ovvero dopo 0.5 per l'impresa 1 e prima di 0.5 per l'impresa 2.

Gli incentivi delle aziende a praticare prezzi più bassi o a addebitare prezzi elevati dipenderanno dall'intensità della concorrenza sul mercato, catturata dal valore del costo di trasporto t : maggiore è il valore di t , minore è la pressione competitiva tra le aziende e più probabile è che colludano e addebitino prezzi elevati.

Un'impresa può riuscire ad accaparrarsi i clienti più vicini all'altra per un'ampia gamma di valori dei parametri (w, t, k) , per i quali gli intervalli disgiunti non sono redditizi per il DB.

Il gioco si struttura come segue: il DB mette all'asta due diverse partizioni delle informazioni.; ciascuna partizione specifica la posizione precisa di ciascun segmento di consumatori sulla linea di Hotelling. Queste partizioni, o strutture di informazioni, possono anche essere diverse per l'impresa 1 e l'impresa 2. Trovare le partizioni ottimali è un problema di ottimizzazione complesso.

Nel modello sono stati calcolati gli intervalli disgiunti ottimali in due casi: quando il DB vende i dati a solo un'azienda e quando li vende a entrambi.

Nel primo caso, il DB vende una partizione che identifica i consumatori vicini all'azienda fino a un punto soglia e lascia i consumatori non identificati nel segmento rimanente, come accennato precedentemente.

Il numero di segmenti in cui i consumatori sono identificati nella struttura ottimale delle informazioni, dipende dal fatto che il DB venda informazioni a una o a entrambe le aziende.

Il data broker combina segmenti della domanda dei consumatori che consentono alle imprese di discriminare il prezzo di terzo grado nei confronti dei consumatori. Si dimostra che il data broker vende informazioni sui consumatori con la più alta disponibilità a pagare e mantiene non identificati i consumatori con bassa disponibilità. Le informazioni sui consumatori a bassa "valutazione" non vengono vendute alle imprese e questi consumatori rimangono non identificati. Facendo ciò, il DB consente alle imprese di estrarre il surplus dai consumatori più preziosi minimizzando la pressione competitiva delle informazioni.

Di conseguenza, si potrebbe dedurre che le imprese che comprano dati dal DB riescono ad estrarre più surplus da un consumatore più disposto a pagare per un certo prodotto (più vicino nella linea di Hotelling) rispetto ad un altro; ai consumatori "lontani", invece, le imprese non hanno incentivo ad addebitare un prezzo più alto. Come si può prevedere anche intuitivamente, non avere informazioni su quei segmenti, non permetterebbe al DB di identificarli e quindi alle imprese di estrarre tutto il surplus.

Il DB sceglie strategicamente di trattenere informazioni sulla domanda dei consumatori per attenuare la concorrenza tra le imprese. Questi risultati si verificano anche sotto discriminazione di prezzo di primo grado, cioè nel caso limite in cui le informazioni sono perfette.

CONCLUSIONI

1. Il data broker non vende informazioni

Le aziende non hanno informazioni sui consumatori, ci si trova nel modello base di Hotelling. All'equilibrio, l'azienda θ imposta $p_\theta = t$ e profitti $\pi_\theta = \frac{1}{2}$

2. Il data broker vende informazioni a una sola azienda

Si prenda in considerazione il caso in cui solo l'impresa 1 abbia acquistato i dati, simmetricamente varrà lo stesso per 2. 1 può distinguere $j + 1$ segmenti di consumatori, con j intero inferiore a k . L'impresa 1 discrimina il prezzo impostando un prezzo per ogni segmento p_{1i} . L'impresa 2, invece, non avendo informazioni sui consumatori, fissa un prezzo uniforme p_2 .

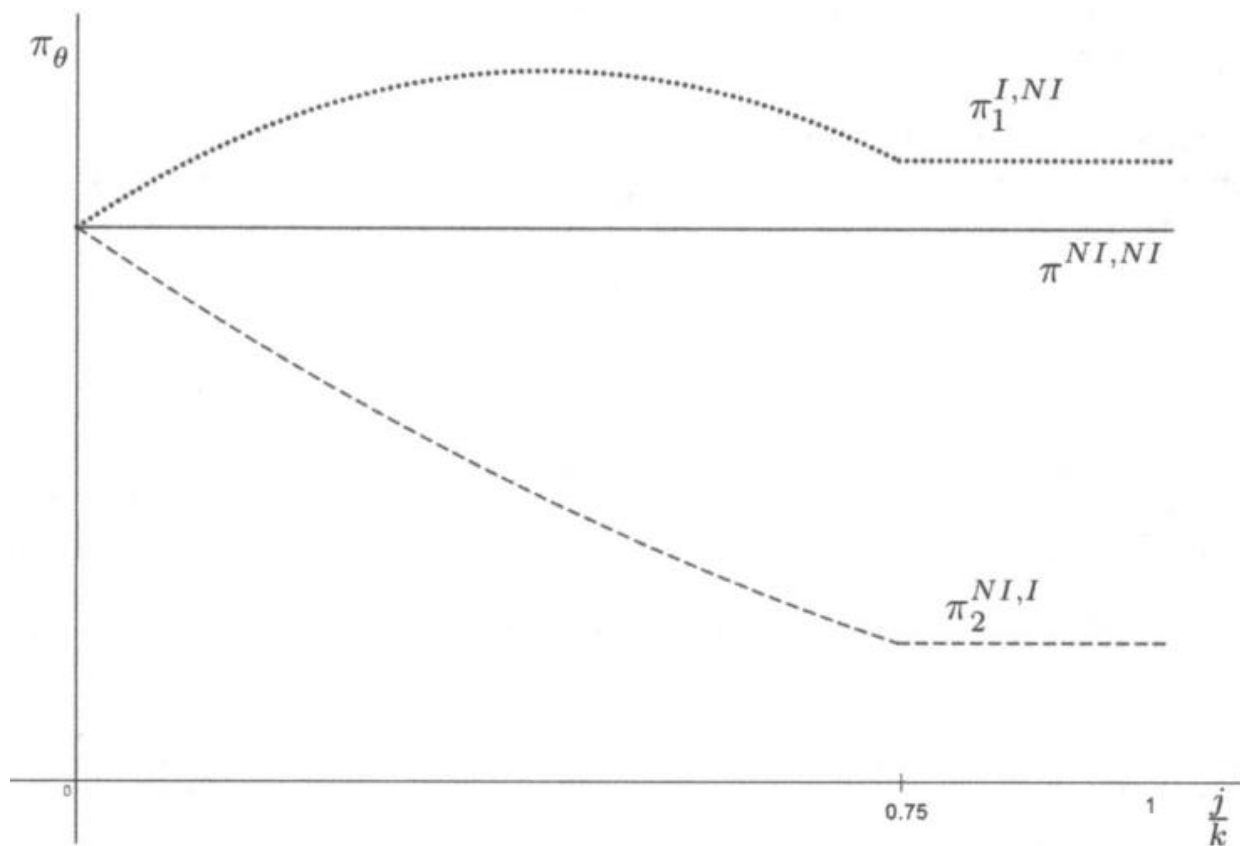


FIGURA 10: PROFITTI NEL CASO 2

In figura, si indica con $\pi_1^{I,NI}$ il profitto dell'impresa 1 se è l'unica ad aver comprato i dati, con $\pi^{NI,NI}$ il profitto di ciascuna delle due imprese nel caso in cui nessuna abbia comprato i dati, e $\pi_2^{NI,I}$ il profitto dell'impresa 2 quando è l'unica a non aver acquistato i dati. Il profitto dell'impresa che possiede i dati è, ovviamente, maggiore in quanto, discriminando, riesce ad estrarre un surplus maggiore ai consumatori con più alta disponibilità a pagare. Inoltre, più informazioni aumentano i profitti dell'impresa 1 solo quando l'effetto di estrazione del surplus domina sull'effetto della concorrenza; infatti, per un valore minimo soglia di j (ovvero da quando 1 ha acquistato una quantità sufficiente di dati), i profitti raggiungono un massimo e poi diminuiscono. Da questo punto, più informazioni portano a una maggiore concorrenza, che domina l'estrazione del surplus del consumatore e riduce quindi, anche, i profitti di 1.

3. Il data broker vende informazioni ad entrambe le aziende

L'impresa 1 identifica j segmenti, e l'impresa 2, rispettivamente, i simmetrici; tuttavia, non vengono acquistati dati su una porzione di segmento al centro in cui le imprese competono. Ciò fa diminuire i profitti come mostrato di seguito.

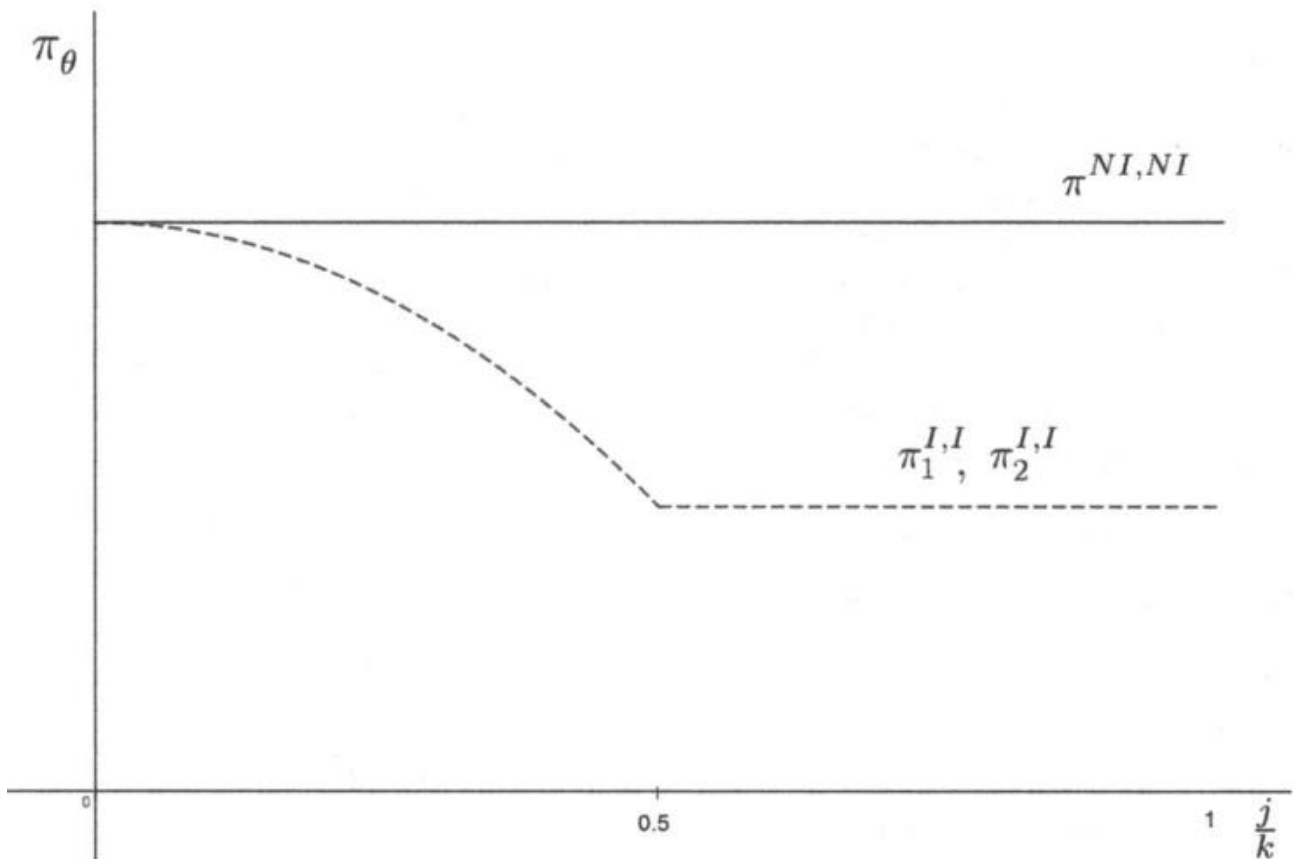


FIGURA 11: PROFITTI NEL CASO 3

Con $\pi_1^{I,I}, \pi_2^{I,I}$ si indicano i profitti di 1 e 2 quando, entrambe, acquistano dati dal DB. I prezzi p_1 e p_2 impostati dalle aziende nel segmento in cui competono, e non discriminano, decrescono in j (j indica il grado di "qualità" di informazione che le aziende acquistano, graficamente corrisponde con un maggior numero di segmenti coperti; $j \in [0, k]$). Anche i prezzi per i consumatori identificati $p_{\theta i}$ decrescono in j , da qui, si può dedurre che più informazioni aumentano la concorrenza tra le aziende, il che riduce i prezzi che esse fissano. La Figura 11 illustra questo effetto quando j aumenta.

Sull'asse orizzontale $\frac{j}{k}$ indica il rapporto tra consumatori identificati e quelli non identificati; infatti, se $j = 0$, $\pi_\theta = \frac{1}{2}$, si sfocia nel modello base di Hotelling (no informazioni), se $j = k$, entrambe le imprese hanno identificato tutti i consumatori e l'effetto competizione ha il predominio. In questo ultimo caso, i profitti delle imprese informate sono minori di quelli che avrebbero senza informazioni; più precisamente, ciò accade già per $j \geq \frac{1}{2}$.

Ulteriori implicazioni delle strategie dei DB sono legate alla politica della concorrenza:

- L'articolo dimostra che è redditizio vendere informazioni anche a un sottoinsieme di imprese presenti sul mercato, escludendo così determinati concorrenti dall'accesso alle informazioni.

Recentemente, Facebook ha offerto ad aziende come Netflix, Lyft o Airbnb un accesso speciale ai dati, mentre ha negato il suo accesso ad altre aziende come Vine. L'accesso ai dati è quindi un fattore critico che influisce sull'intensità della concorrenza di un settore.

- In secondo luogo, con lo sviluppo del big data, machine learning e della AI, i data broker possono segmentare meglio la domanda dei consumatori. Pertanto, poiché si passerebbe dalla discriminazione di prezzo di terzo grado alla discriminazione di primo grado, il benessere dei consumatori diminuirebbe ancora di più.

Il secondo modello (Gu, et al., 2021) è costituito da due mercati interconnessi; è simile al precedente, ma vi è la presenza di concorrenza anche nel mercato a monte; infatti, due DB (DB1 e DB2) competono e hanno informazioni sui consumatori che comprano i prodotti in un mercato a valle. Nel mercato a valle, le aziende sono differenziate in modo orizzontale e si parla del concetto di sub e super-additività, meglio spiegato nel modello successivo. Vi è la presenza di una nuova componente: i DB possono condividere i propri dati e venderne un singolo set. In questo caso, agiscono congiuntamente come venditore unico di dati e fanno un'offerta "prendere o lasciare (TILI)" a uno specifico acquirente.

Indicando con k la generica impresa, il profitto congiunto risulta essere:

$$\pi_k = \begin{cases} 0 & \text{se le aziende a valle non comprano i dati congiunti} \\ s_k(p_{12} - c_{db}) & \text{altrimenti;} \end{cases}$$

Con p_{12} si identifica il prezzo fissato congiuntamente dai due DB, $c_{db} > 0$ è il costo di fusione sostenuto dai broker di dati a monte e $s_k \in [0,1]$ è la quota del profitto comune data da una regola di ripartizione fissata esogenamente. Conseguentemente e analogamente ai casi precedenti, si ricavano i profitti delle imprese a valle, a seconda che comprino o meno i dati dai DB.

Il timing del gioco è il seguente: nella prima fase, i due DB decidono contemporaneamente e indipendentemente se condividere o meno i propri dati. La condivisione dei dati avviene se, e solo se, entrambi i DB scelgono di condividere i dati. Nella seconda fase, i data broker fissano congiuntamente, se nella fase precedente hanno deciso di vendere un unico data set, o indipendentemente, il/i prezzo/i per il/i set di dati. Nella terza, l'acquirente decide se acquistare o meno il/i set di dati offerto/i. Per risolvere il gioco, è stato identificato l'equilibrio di Nash perfetto nei sottogiochi (SPNE) tramite backward induction.

Si definisce: $f_k > 0$, l'extra surplus che l'acquirente può generare utilizzando i dati di proprietà del DB_k . La funzione valore, " f_k " può essere interpretata come la valutazione monetaria del set di dati dal punto di vista dell'acquirente. Senza perdita di generalità, si suppone $f_1 \leq f_2$; C_b è il costo di fusione dei dati a valle da parte dell'azienda acquirente.

Si ottiene, dunque, la seguente struttura dei dati:

- super-additività a valle, se $f_{12} - C_b \geq f_1 + f_2$
- sub-additività a valle, se $f_2 \leq f_{12} - C_b < f_1 + f_2$, e infine
- sub-additività estrema a valle, se $f_{12} - C_b < f_2$.

Il sottogioco in cui entrambi i DB decidono di vendere un dataset congiunto è descritto come segue: i DB agiscono come fornitore esclusivo dell'impresa a. Poiché essi fanno congiuntamente un'offerta "TILI" all'acquirente, se la struttura dei dati è super o sub-additiva a monte, il profitto totale che i DB possono ottenere è $f_{12} - C_{db}$. Se la struttura dei dati è sub-additiva estrema a monte, i data broker non procedono alla fusione dei set di dati e vendono semplicemente quello di maggior valore all'acquirente, ovvero f_2 , ottenendo congiuntamente il suo valore. In definitiva, per $k = 1,2$

$\pi_k^s = s_k * \max\{f_{12} - c_{db}, f_2\}$, dove π_k^s indica il profitto del DB k-esimo sotto data sharing.

Affinché la condivisione dei dati possa portare a un equilibrio perfetto nei sottogiochi (SPNE), il profitto congiunto deve essere almeno pari a quello di concorrenza, cioè $\pi_1^s + \pi_2^s \geq \pi_1^c + \pi_2^c$; altrimenti, la condivisione non può essere una risposta ottimale reciproca nel primo stadio del gioco.

È più naturale pensare $C_b > C_{db}$, ovvero l'acquirente è meno efficiente dei DB nell'unire i set di dati; ad esempio, una compagnia assicurativa desidera accedere a diverse caratteristiche dei potenziali clienti per il credit scoring e la profilazione. Per il punteggio del credito, i dati relativi al reddito e alla ricchezza possono essere ottenuti tramite un DB di dati finanziari; per ottenere dati sulla profilazione, può essere utilizzata la cronologia di navigazione, facilmente ottenibile da un DB specializzato in marketing. I due DB gestiscono abitualmente tali informazioni, a differenza della compagnia assicurativa, per la quale unire e "pulire" database separati rappresenta un compito notevolmente più complesso. Di conseguenza, i DB si trovano collettivamente in condizioni migliori quando condividono i dati poiché ciò aiuta a internalizzare le inefficienze a valle ed evitare la concorrenza quando i loro set di dati si sovrappongono.

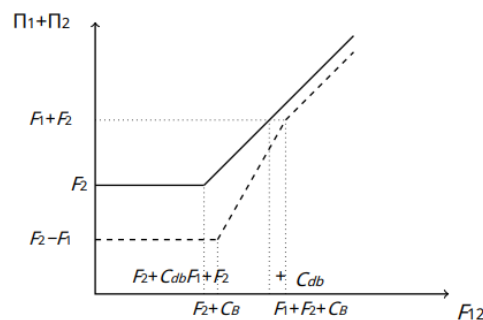


FIGURA 12: $C_b > C_{db}$

Nella figura 13, in cui $C_b > C_{db}$, sulle ordinate, sono rappresentati i profitti congiunti dei data broker derivanti dalla condivisione (linea continua) e dalle vendite individuali (linea tratteggiata); in questo caso i profitti congiunti generati dalla condivisione saranno sempre al di sopra di quelli generati dalla vendita individuale.

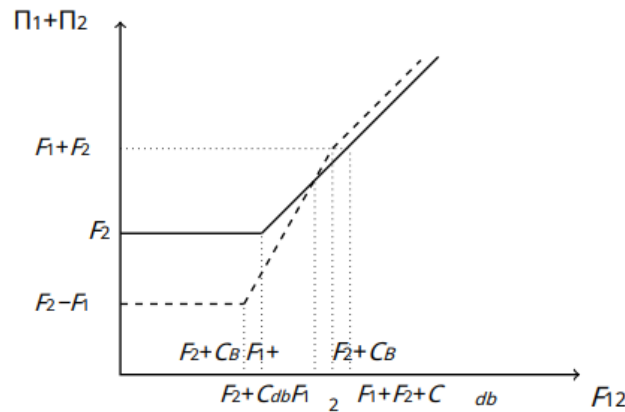


FIGURA 13: L'acquirente è più efficiente, $C_{db} - \frac{f_1}{2} \leq C_b < C_{db}$

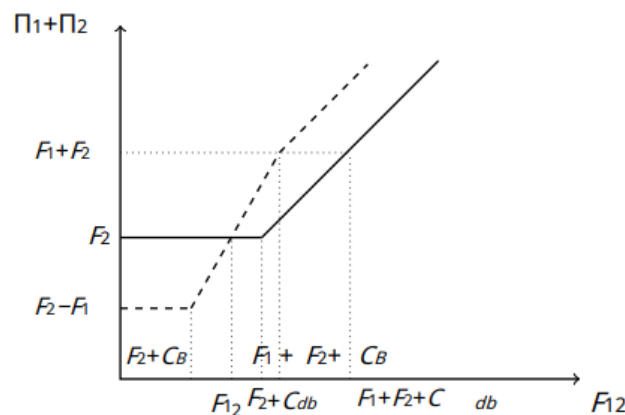


FIGURA 14: L'acquirente è molto più efficiente, $C_B < C_{db} - \frac{f_1}{2}$

Nelle figure 14 e 15, i DB non sono abbastanza efficienti nell'unire i dataset.

Un aspetto distintivo del settore dei DB, come evidenziato dal rapporto della Federal Trade Commission (2014), è lo scambio e il commercio di dati tra broker e come questo si relaziona alle particolari proprietà dei dati, rispetto ad altri prodotti (valore contestuale, costi di fusione, complementarità).

Le conclusioni del modello in esame sono compatibili con un mercato dei dati in cui i data broker interagiscono ripetutamente per fornire acquirenti in diversi sottomercati. Gli incentivi per la condivisione dei dati sono strettamente legati alla natura stessa dei dati. In particolare, si trova che la condivisione può emergere per due motivi:

1. i DB possono attenuare la concorrenza quando i dati presentano una qualche forma di sostituibilità;
2. ciò permette ai data broker di internalizzare le inefficienze a valle, poiché gli acquirenti potrebbero essere meno efficienti dei data broker nel fondere più set di dati.

A sua volta, si identifica un trade-off tra gli effetti positivi dell'internalizzazione dei costi, in linea con lo spirito della strategia dei dati dell'UE (2020) e gli effetti negativi della condivisione dei dati legati alla riduzione della concorrenza; in particolare, l'importanza delle strutture dei dati sub o super-additivi e dei costi di fusione dei dati.

La collaborazione può andare oltre le situazioni di creazione di valore e può attenuare la concorrenza tra i data broker a scapito dei loro clienti; per contrastare il potere di mercato dei DB, l'autore suggerisce che un'azienda potrebbe preferire acquistare dati di "qualità inferiore" (ad esempio, sub-additivi, con informazioni sovrapposte) per aumentare la competizione tra i data broker; l'azienda può così trattenere parte del surplus prodotto attraverso i dati. Inoltre, le inefficienze dei costi a valle potrebbero rivelarsi un vantaggio poiché la competizione porta i DB a concedere uno sconto all'azienda.

Il terzo lavoro preso in esame (Abrardi, et al., 2024) è un gioco dinamico che presenta concetti simili al precedente, in quanto riprende lo schema a monte (con due DB), a valle con le imprese che competono con prodotti sostituti; le imprese a valle possono acquistare i dati dai DB: da uno solo di essi, da entrambi o non comprarne, e possono offrire prezzi discriminati ai consumatori. In questo modello viene approfondito il concetto di sub e super-additività.

La struttura del modello è la seguente: nel mercato downstream, si prende in considerazione una città circolare con libero accesso, in cui un numero $n > 2$ di imprese entra nel mercato. Le imprese sono indicate con i , con i intero che varia da 1 a $n-1$, e vendono prodotti concorrenti ai consumatori. L'ingresso delle imprese è sequenziale per evitare problemi di coordinamento. Le imprese entrano nel mercato scegliendo posizioni equidistanti, in modo che una generica impresa i sia situata in i/n . I costi marginali delle imprese sono normalizzati a zero, mentre il costo fisso di ingresso è F , che può essere interpretato come il costo della digitalizzazione; ad esempio, l'investimento necessario per aprire un negozio e-commerce.

I consumatori sono distribuiti uniformemente sulla circonferenza, con massa totale normalizzata a 1. La posizione di un generico consumatore è $x \in [0, 1)$ e viene considerata crescente in senso antiorario; ognuno di loro acquista al massimo un'unità del prodotto. L'utilità lorda derivante dal consumo è v , e il costo di trasporto per consumatori è lineare, ed è definito con t .

Nel mercato a monte, DB1 e DB2 hanno set di dati contenenti informazioni sui clienti che consentono alle imprese di identificare i consumatori con una certa probabilità. Il set di dati del DB1 contiene informazioni che danno alle imprese una probabilità $\alpha \in [0, 1]$ che consente loro di identificare i consumatori e, su coloro che sono stati identificati, di operare una discriminazione di primo grado. Invece, il set di dati del DB2 contiene informazioni che danno alle imprese una probabilità $\beta\alpha$, con $\beta \in [0, 1)$, di identificare i consumatori. Il parametro α può essere interpretato come l'accuratezza dei dati, invece, β come il grado di differenziazione verticale tra i DB.

La combinazione dei set di dati dei DB fornisce una ulteriore accuratezza γ , con $\alpha \leq \gamma \leq 1$ su quei consumatori. L'accuratezza dei set di dati combinati γ può essere interpretata come il livello di sinergie tra i due set di dati. Se un'impresa decide di acquistare i dati da entrambi i DB, l'accuratezza del dataset combinato può essere inferiore (sub-modular, cioè la capacità predittiva dei dataset combinati è inferiore alla somma delle capacità predittive dei singoli) o superiore (super-modular) alla somma delle accuratèzze individuali. La sub-modularità potrebbe derivare, ad esempio, da un sovrapporsi parziale tra i dataset, o da rendimenti decrescenti.

I DB possono vendere partizioni dei loro set di dati alle imprese. Una partizione, come per il modello precedente, fornisce informazioni sulla posizione dei consumatori in un segmento di mercato. In un modello basato sul mercato di Hotelling come questo caso, a differenza del precedente, le partizioni dati non possono contenere intervalli disgiunti e, si dimostra che le partizioni di equilibrio devono contenere le posizioni delle imprese (sia per discriminazioni di primo che di terzo grado); in questo caso le partizioni offerte a ciascuna impresa sono centrate rispetto alla posizione di quest'ultima. Ogni DB_k , con $k \in \{1, 2\}$, sceglie la dimensione della partizione $d_{i,k} \in [0, \frac{1}{n}]$ da offrire all'impresa i . Quindi, la partizione venduta dal DB1 consente alla generica impresa i di identificare un segmento di dimensione $d_{i,1}$ di consumatori con probabilità α , mentre la partizione venduta dal DB2 consente di identificare $d_{i,2}$ consumatori con probabilità $\beta\alpha$. I consumatori situati all'interno sia $d_{i,1}$ che $d_{i,2}$, seguendo le definizioni precedenti, sono identificati con probabilità γ .

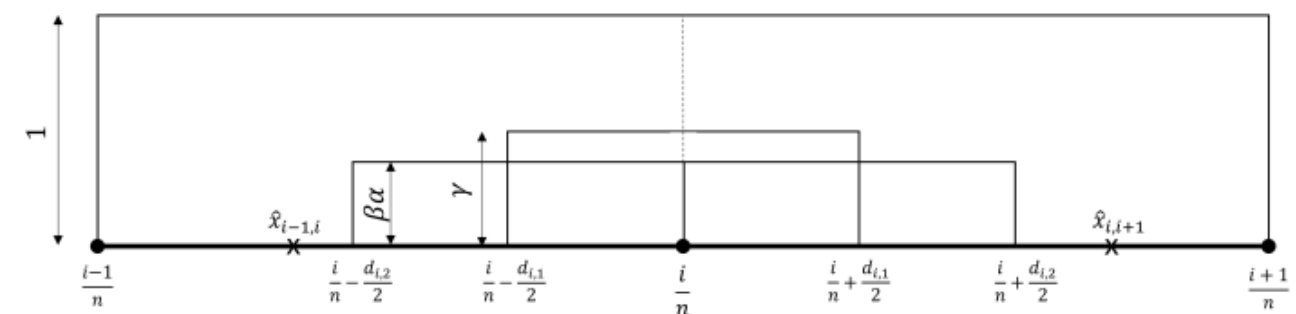


FIGURA 15: Quota di mercato dell'impresa i -esima se compra sia dal DB1 Che dal DB2 ($d_{i,1} < d_{i,2}$)

Se γ sia maggiore o minore di α , dipende dalla "super" o "sub-modularità".

Si assume che ogni consumatore osservi solo un prezzo da una determinata impresa e che i consumatori preferiscano i prezzi personalizzati rispetto ai prezzi standard quando sono indifferenti; di conseguenza, indicando con p_i il prezzo che il consumatore paga (indipendentemente se identificato o meno), la sua utilità sarà:

$$U(x, i) = v - p_i(x) - t * D(x, i)$$

con $D(x, i)$ si indica l'arco più breve tra la posizione del consumatore x e la posizione dell'impresa i , $\frac{i}{n}$.

I due DB vendono contemporaneamente partizioni di dati attraverso offerte non rinegoziabili "take it or leave it", prendere o lasciare. Con $w_{i,k}$ il prezzo offerto dal DB_k per la partizione di dati $d_{i,k}$.

Il timing del modello è il seguente:

1. Le imprese entrano nel mercato e pagano il costo fisso F .
2. Ciascun $DB_k \in \{1, 2\}$ sceglie una partizione $d_{i,k}$ per ciascuna impresa e la offre a un prezzo $w_{i,k}$.
3. Ciascuna impresa che è entrata nel mercato decide se accettare o declinare le offerte dei DB.
4. Le imprese stabiliscono prezzi standard per i consumatori non identificati, ovvero i consumatori situati al di fuori della partizione di dati, e per i consumatori situati all'interno della partizione nel caso in cui non riescano a identificarli.
5. Le imprese che hanno ottenuto una partizione stabiliscono prezzi discriminati per i consumatori identificati.

Il modello è stato risolto tramite backward induction, quindi partendo dalla risoluzione del mercato a valle.

1. Tutte le imprese acquistano da entrambi i DB

Quando tutte le imprese acquistano dati sia dal DB1 che dal DB2, i prezzi standard e personalizzati in equilibrio diminuiscono in α , γ , $d_{i,1}$, $d_{i,2}$ e diminuiscono debolmente in β , per ogni impresa.

Con l'aumentare della quantità di dati o della loro accuratezza, o con un rafforzamento delle sinergie, un'impresa identifica una quota maggiore di consumatori. Di conseguenza, il prezzo standard di equilibrio viene offerto ai consumatori che, in media, sono più lontani dalla posizione dell'impresa, cioè coloro che hanno una minore propensione a pagare per il prodotto. A sua volta, la riduzione dei prezzi non discriminati, esercita una pressione al ribasso sui prezzi discriminati, così come sui profitti delle imprese; questo è l'effetto concorrenza già citato per il modello precedente.

Allo stesso tempo, i dati consentono alle imprese di offrire prezzi discriminati ai consumatori con più alta disponibilità a pagare (nel modello, i più vicini) che estraggono in

modo più efficace il surplus del consumatore e aumentano i profitti dell'impresa. Nel complesso, si generano effetti finali contrastanti dovuti all'effetto concorrenza e all'effetto di estrazione del surplus.

L'effetto di β sui prezzi all'equilibrio è negativo, ma solo quando $d_{i,2} \geq d_{i,1}$, perché alcuni consumatori sono identificati solo attraverso le partizioni di DB2 con un'accuratezza $\beta\alpha \neq 0$, e quindi preferibile rispetto a non ottenere dati su quei consumatori. Se invece $d_{i,2} < d_{i,1}$, nessun consumatore è identificato solo dalla partizione di DB2, e β non influenzerà i prezzi all'equilibrio.

2. Sottogioco in cui un'impresa acquista solo da un DB

In un generico sottogioco in cui l'impresa i acquista da solo un DB_k i dati e le altre imprese da entrambi, visto che l'impresa i acquista meno dati rispetto alle altre imprese, i suoi prezzi di equilibrio sono più alti e, per la complementarità delle strategie di prezzo, anche i prezzi delle altre imprese sono più alti. In particolare, il prezzo standard dell'impresa i diminuisce nel totale dei consumatori identificati.

Nello specifico sottogioco in cui l'impresa i acquista dati solo da DB2, mentre tutte le altre imprese acquistano dati da entrambi i DB, i prezzi standard e discriminati di tutte le imprese in equilibrio sono più alti rispetto ai prezzi nel sottogioco in cui tutte le imprese acquistano da entrambi i DB. Inoltre, i prezzi standard di equilibrio sono più alti nel sottogioco in cui l'impresa i acquista da DB2 rispetto a quando acquista da DB1 se e solo se $\beta d_{i,2} < d_{i,1}$. Questo perché, se l'impresa i acquista da DB1, identifica $\alpha d_{i,1}$ consumatori, mentre se acquista da DB2 identifica $\alpha\beta d_{i,2}$ consumatori. Quindi, se $\beta d_{i,2} < d_{i,1}$, cioè se i consumatori identificati quando solo i acquista da DB2 è minore del numero di consumatori identificati quando acquista solo da DB1, fissa un prezzo standard più alto in quel sottogioco.

Di seguito l'analisi del caso in cui i dataset non presentano sinergie:

Per quanto riguarda il mercato a monte, viene preso come riferimento il caso in cui il dataset di un data broker sia contenuto nell'altro, ad esempio, DB1 possiede i dati anagrafici e lavorativi dei consumatori, invece, sugli stessi consumatori, DB2 possiede solo quelli anagrafici; ciò significa che per una qualsiasi impresa non è razionali acquistarli entrambi. Se il dataset di DB2 è compreso in quello di DB1, si rientra nel caso in cui $\gamma = \alpha$. Questo è il caso in cui la competizione è più forte, in quanto, DB2, non avendo dati esclusivi, ha bisogno di abbassare i prezzi per poter competere con DB1.

I DB competono sia nei prezzi che nelle dimensioni delle partizioni che vendono alle imprese a valle. Nel caso di riferimento, il set di dati del DB2 contiene informazioni con una precisione inferiore rispetto al set di dati del DB1 per ogni consumatore considerato. Di conseguenza, in equilibrio, solo il DB1 vende partizioni di dati nel mercato a valle, mentre il DB2 non realizza profitti. Tuttavia, il DB2 esercita pressione competitiva sul DB1, limitando la sua capacità di estrarre profitti dalle imprese. DB1 fissa il prezzo per i dati $w_{i,k}$

pari alla “propensione al pagamento” delle imprese, che è la differenza nei profitti delle imprese tra l'acquisto di dati da DB1 o da DB2.

- In equilibrio, sia DB1 che DB2 offrono partizioni di dimensioni uguali alle imprese che entrano nel mercato, cioè $d_{i,1} = d_{i,2} < \frac{1}{n}, \forall i \in \{0, \dots, n-1\}$. Oltre a offrire stesse dimensioni delle partizioni tra DB, essi offrono anche esattamente le stesse partizioni alle imprese, che vengono definite con d^* ; il valore di d^* è decrescente in α e crescente in β . Tutte le imprese acquistano da DB1 e i profitti di equilibrio di DB1 aumentano in α e diminuiscono in β . Gli effetti di α e β su d^* sono determinati dall'interazione tra l'effetto di estrazione del surplus e l'effetto concorrenziale.

Un aumento di α , pur ingrandendo entrambi gli effetti, ha un impatto maggiore sull'effetto concorrenziale, a causa della complementarità tra le strategie di prezzo delle imprese. Per mitigare l'effetto concorrenziale, che influisce negativamente sui profitti delle imprese, i DB riducono la dimensione delle partizioni di equilibrio. Al contrario, un aumento di β favorisce le imprese se acquistano dati da DB2. Quindi, il DB1 aumenta la dimensione della partizione per ridurre i profitti delle imprese nel caso in cui acquistino dati da DB2. In questo modo, aumenta la loro propensione a pagare per i suoi dati. Guardando ai profitti di equilibrio del DB1, aumentano con l'accuratezza delle informazioni (α) poiché i dati diventano più preziosi per le imprese, e diminuiscono man mano che la pressione competitiva da parte del DB2 (β) si fa più forte.

Quando vi è sinergia tra i dataset, invece:

Si supponga, ad esempio, che il DB1 abbia informazioni sul genere, reddito e occupazione dei consumatori, mentre DB2 sull'occupazione e sull'indirizzo; in questo caso vi è una parziale sovrapposizione tra parti dei dataset; tuttavia, potrebbe essere razionale acquistare partizioni di entrambi.

A seconda del livello di sinergie del dataset combinato, si possono verificare i due casi citati precedentemente:

1. se $\gamma < \alpha + \beta\alpha$, i due dataset sono sub-modulari; ad esempio, contengono alcune informazioni sovrapposte;
2. se invece $\gamma \geq \alpha + \beta\alpha$, i due dataset sono super-modulari, cioè ci sono complementarità tra le informazioni nei due dataset.

L'obiettivo dell'articolo è quello di evidenziare come la qualità delle informazioni (precisione) dei DB e le sinergie tra i loro dataset influenzino la concorrenza e le loro strategie. Si distinguono due dimensioni dei dati:

- Qualitativa (l'accuratezza dei dati) che riflette l'efficacia delle informazioni raccolte dai DB su ogni consumatore per fare inferenze sulle sue preferenze.

- Quantitativa (partizioni dei dati) che riflette la copertura dei dataset venduti dai DB in termini di consumatori. Ad esempio, una partizione può essere composta da dati su consumatori che risiedono in una specifica regione geografica.

Nel modello in esame, due DB decidono il prezzo e la dimensione delle partizioni, che vendono a ciascuna impresa in un mercato a valle con prodotto differenziato (Salop 1979), dove i dati consentono alle imprese di operare una discriminazione del prezzo di primo grado sui consumatori identificati. La discriminazione è imperfetta, cioè un generico consumatore può essere identificato con una probabilità inferiore a uno da ciascuno dei due DB, che hanno diversi livelli di accuratezza.

- Il principale risultato è che i DB, coordinandosi sul prezzo dei dati e limitando opportunamente la dimensione delle partizioni, possono rendere il valore dei dataset combinati più alto della somma dei valori individuali (dataset super-additivi) anche se i dati sono sub-modulari (a condizione che i rendimenti non siano troppo deboli), inducendo così le imprese ad acquistare entrambi i dataset. Infatti, il dataset congiunto consente alle imprese di ottenere una maggiore accuratezza attraverso le sinergie, aumentando la loro disponibilità a pagare per i dati. Questo intensifica la concorrenza tra le imprese, riducendo i loro profitti e la loro disponibilità a pagare per i dati.
- Vendendo partizioni più piccole, i DB possono mitigare la concorrenza a monte, indurre la vendita dei dataset combinati e far pagare un prezzo più alto per i dati, riducendo così l'ingresso a valle e causando danni al surplus dei consumatori.
- Il modello dimostra che l'additività è una caratteristica endogena dei dati e che può essere influenzata dai DB attraverso la dimensione delle partizioni. Inoltre, si mette in evidenza come il prezzo dei dati e le strategie di coordinamento dei DB influenzino l'ingresso a valle e i profitti.
- La capacità dei DB di trasformare i dati in informazioni utili per estrarre valore consente di influenzare la concorrenza a valle, con significative implicazioni per il benessere. La concorrenza tra DB e le sinergie dei dati influenzano le strategie e, di conseguenza, gli esiti economici.
- Il risultato principale evidenzia un nuovo canale attraverso il quale i DB possono mitigare la concorrenza che li coinvolge nel mercato a monte, anche se le sinergie dei dati sono deboli. Scegliendo strategicamente il prezzo e la dimensione delle partizioni che vendono alle imprese, i DB possono aumentare endogenamente il valore dei dati, ciò permette loro di coordinare la vendita e di estrarre maggiori profitti dalle imprese a valle. Così, anche in assenza di collusione, i DB possono comportarsi strategicamente per ottenere una super-additività nel valore dei dati **anche se i dataset sono sub-modulari**.
- Infine, le sinergie tra DB concorrenti riducono l'ingresso a valle. Ciò può causare danni ai consumatori. Se la concorrenza tra i DB non è sufficientemente forte, sia a

causa di **forti sinergie tra i dataset** che di un **elevato livello di differenziazione verticale**, la riduzione dell'ingresso domina l'effetto della concorrenza a valle.

Il quarto modello (Krämer, et al., 2018) è costituito da un gioco che ha il fine di esaminare l'impatto dei personal data broker sugli esiti di mercato e il surplus dei consumatori.

È utile definire come si sviluppa il mercato inizialmente senza il PDB e successivamente con la presenza del PDB.

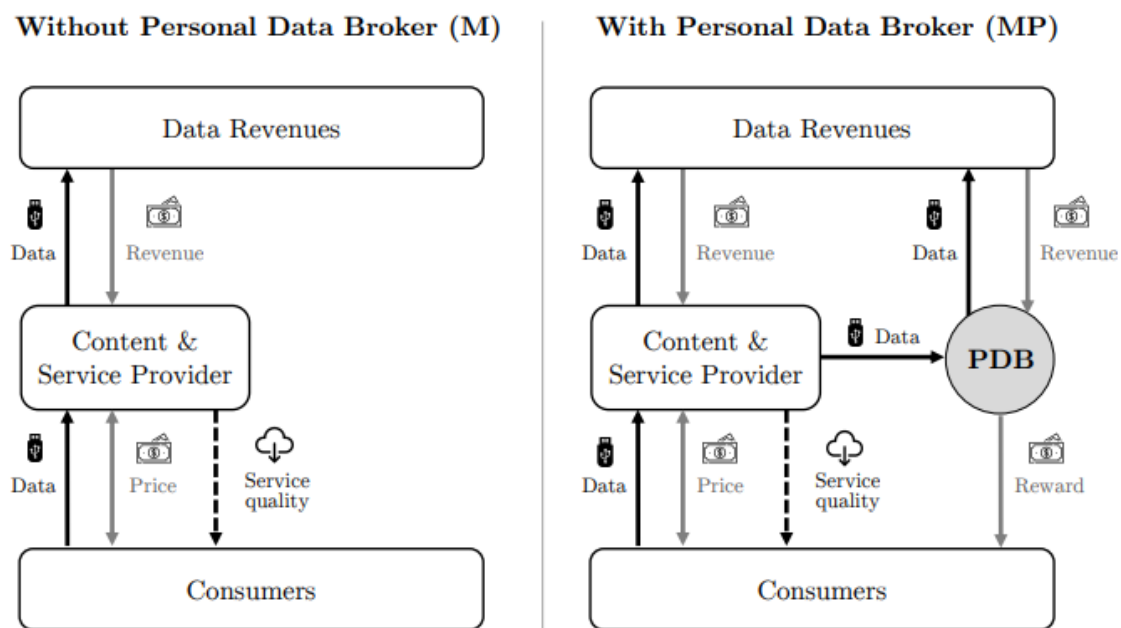


FIGURA 16: struttura del mercato senza PDB (sinistra) e con (destra)

- Nella struttura di sinistra, il CSP ha due flussi di entrate. Può addebitare ai consumatori un prezzo (che può, ovviamente, anche essere zero o negativo) per l'uso del suo servizio, e può generare entrate monetizzando i dati dei consumatori. Il CSP può determinare strategicamente sia la qualità del suo servizio che il prezzo di utilizzo per massimizzare il suo profitto totale. Più alto è il livello di qualità del CSP, o più basso è il suo prezzo, più consumatori utilizzano il servizio e quindi forniscono più dati al CSP.
- Nella configurazione di destra, con un PDB, i consumatori possono inoltre trasferire i loro dati al PDB in cambio di un premio monetario positivo dal PDB. Il PDB può quindi monetizzare questi dati contemporaneamente al CSP competendo per le entrate provenienti dai dati.

Il timing del gioco rappresentato di seguito:

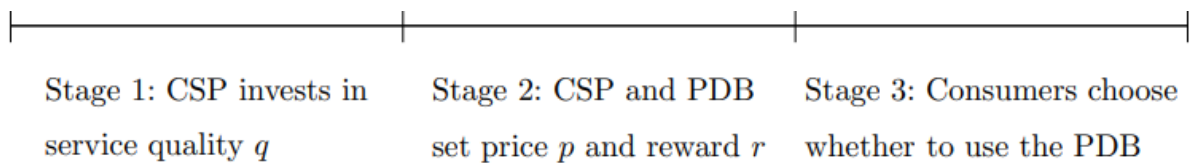


FIGURA 17: *timing del gioco*

Dalla Figura 17 si evince che l'investimento nella qualità del CSP è una decisione a lungo termine, mentre il prezzo del CSP e il premio del PDB possono essere regolati nel breve termine.

Si considera una massa unitaria di consumatori con una domanda totale per il servizio del CSP pari a $x(q, p, r)$, dove q indica la qualità del servizio del CSP, p il prezzo del servizio e r la ricompensa pagata dal PDB, se presente. Per mantenere il modello semplice, si assume che la quantità di dati che può essere raccolta dal CSP sia direttamente proporzionale al livello di utilizzo dei consumatori, e si denotano entrambi con x . Quindi, ci si riferisce a x sia come al livello di utilizzo che alla quantità di dati forniti. Si può pensare a x come al tempo che gli utenti trascorrono presso un CSP, lasciando una traccia di dati. Se i consumatori decidono, invece, di utilizzare il PDB, trasferiscono i loro dati di utilizzo x dal CSP al PDB.

Il modello prende in considerazione il caso in cui un CSP imposta strategicamente la qualità di un servizio offerto ai consumatori (gratuitamente o a pagamento). Utilizzando il servizio, i consumatori generano dati personali, ad esempio, rivelando le proprie preferenze. Un maggiore utilizzo del servizio consente al CSP di raccogliere più dati degli utenti, che a sua volta genera maggiori entrate da dati, come dalla pubblicità. Più entrate possono essere generate dai dati degli utenti, maggiore è l'incentivo del CSP a stimolare l'utilizzo del suo servizio, aumentando la qualità del servizio o riducendo il prezzo del servizio, o entrambi. Con l'introduzione di un PDB, i consumatori possono trasferire i loro dati personali generati durante l'uso del servizio CSP al PDB. Il PDB ricompensa i consumatori per il trasferimento dei dati e può quindi derivare entrate da dati in competizione con il CSP. Sono stati analizzati e confrontati gli esiti di equilibrio in un contesto di mercato con e senza un PDB.

Il PDB genera ricavi solo dai dati e può incentivare i consumatori a trasferire e a generare dati presso il CSP pagando una ricompensa $R(x) = rx$, con $r > 0$, altrimenti i consumatori non avrebbero alcun incentivo a trasferire i dati al PDB. Per effettuare un confronto significativo tra la situazione con e senza un PDB, è importante che il CSP e il PDB abbiano gli stessi strumenti di pricing disponibili per incentivare l'utilizzo del servizio e la generazione di dati; diversamente, i risultati potrebbero essere influenzati dal fatto che l'uso

può essere incentivato in modo diverso dal PDB rispetto al CSP. Nel modello in questione, sia il CSP che il PDB possono stimolare l'uso attraverso prezzi lineari. I consumatori, quindi, prendono la loro decisione di consumo in base al prezzo netto $n = p - r$

Il modello mostra che possono emergere due principali esiti di mercato, a seconda dell'efficienza del PDB nella generazione di entrate da dati.

1. Reddito minimo (**minimum income equilibrium**). Se l'efficienza del PDB è bassa, il PDB offre ai consumatori solo una ricompensa minima, sufficiente unicamente per garantire la partecipazione dei consumatori, denominata equilibrio di reddito minimo. Anche se tutti i consumatori utilizzano il PDB, le ricompense sono basse e il CSP riduce la qualità del suo servizio in risposta alle minori entrate da dati a causa della concorrenza del PDB. Pertanto, in questo caso, il mercato dei dati personali fallisce e il surplus del consumatore risulta essere inferiore rispetto a quello senza il PDB.
2. Solo se il PDB è sufficientemente efficiente, può offrire ricompense positive e significative ai consumatori. Questo risultato è stato denominato di mercato equilibrio di reddito positivo (**positive income equilibrium**). Tuttavia, anche nell'equilibrio di reddito positivo, i consumatori potrebbero comunque trovarsi in una situazione peggiore rispetto a quella senza il PDB. Ciò avviene perché il CSP potrebbe abbassare comunque la qualità del suo servizio quando il PDB eroga una ricompensa elevata, poiché i consumatori hanno ora un ulteriore incentivo ad utilizzare il servizio del CSP e a generare comunque dati; allo stesso tempo, il CSP è in grado di aumentare il prezzo del servizio poiché i consumatori ricevono ricompense positive. In sintesi, un PDB migliora il benessere dei consumatori solo se l'efficienza del PDB è così elevata da far sì che le entrate complessive del settore dei dati siano superiori rispetto a quelle senza il PDB.

In conclusione, sono stati analizzati i seguenti tre effetti strategici:

- Effetto competizione
 - Nell'equilibrio del reddito minimo e nell'equilibrio del reddito positivo, il CSP ha un incentivo inferiore a investire nella qualità del servizio rispetto a quanto avviene senza il PDB, a causa della diminuzione dei ricavi dati del CSP dovuta alla concorrenza del PDB. Lasciando tutte le altre variabili inalterate, più è forte la riduzione dei ricavi dati dovuta al PDB, più bassa è la qualità di equilibrio del CSP. Nell'equilibrio del reddito minimo, il prezzo netto per i consumatori è lo stesso di quello senza il PDB. Pertanto, è solo la riduzione strategica della qualità del CSP (a causa dell'effetto della concorrenza) che determina le differenze nei risultati di mercato.
- Effetto spostamento (displacement effect)

- Nell'equilibrio del reddito positivo, non solo il CSP incentiva i consumatori a generare dati (impostando qualità e prezzo), ma anche il PDB, stabilendo una ricompensa positiva. Tuttavia, la ricompensa del PDB r e la qualità del CSP q sono sostituti strategici, poiché un aumento della qualità comporta una riduzione della ricompensa: $\partial r / \partial q < 0$. Pertanto, nell'equilibrio di reddito positivo, il CSP ha meno incentivi a impostare un livello di qualità più elevato per stimolare la generazione di dati dei consumatori, perché prevede che l'aumento della qualità comporterebbe una riduzione strategica della ricompensa del PDB, che a sua volta non stimolerebbe sufficientemente la generazione di dati dei consumatori.
- Effetto appropriazione
 - Man mano che il PDB diventa più efficiente e paga ricompense più alte, il CSP aumenta strategicamente il suo prezzo per appropriarsi di parte del surplus aggiuntivo dei consumatori generato dal PDB. Formalmente, ciò può essere visto dalla ricompensa del PDB e dal prezzo del CSP che sono complementi strategici, cioè in presenza del PDB $\partial p(q)/\partial r > 0$ e $\partial r(q)/\partial p > 0$. Prevedendo che il guadagno di efficienza del PDB possa essere appropriato attraverso il suo prezzo nello stadio 2 del gioco, il CSP ha un incentivo aggiuntivo ad aumentare la sua qualità nello stadio 1 all'aumentare dell'efficienza del PDB. Questo stimola ulteriormente la generazione di dati da parte dei consumatori che vengono poi ricompensati.

Conseguentemente a queste conclusioni, si può dedurre che molti PDB sono attualmente bloccati nell'equilibrio di reddito minimo, poiché le ricompense che gli utenti possono ricevere sono spesso molto limitate, nonostante vi sia il sostegno regolamentare e una crescente, ma forse non sufficiente, consapevolezza tra i consumatori che i loro dati personali abbiano un valore. Infatti, ci sono anche diversi esempi di PDB che sono falliti, sollevando il dubbio se "tali mercati per i dati personali potrebbero mai avere successo".

Anche con i diritti alla portabilità dei dati per i consumatori, i Personal Data Broker hanno scarse probabilità di ottenere una forte presa sui mercati digitali, e quindi i mercati dei dati personali sono inclini a fallire. Inoltre, è stato rilevato che l'emergere dei PDB potrebbe alla fine danneggiare i consumatori.

Un esempio della difficile sopravvivenza dei PDB è quello di Datacoup: fondata nel 2012 e con sede a New York; gli utenti di Datacoup potevano collegare i loro account social e finanziari alla piattaforma e selezionare quali dati condividere. In cambio, ricevevano un compenso mensile, che durante il periodo beta era fino a 8 dollari al mese. Ha ottenuto finanziamenti totali per circa 1.41 milioni di dollari. Nel novembre 2019, tuttavia, Datacoup ha annunciato la chiusura delle operazioni e la cancellazione di tutti i dati memorizzati sui loro server; è stata acquisita da ODE Infinity il 1° novembre 2022 (PitchBook, 2024).

Diversamente, Datum ha preferito una strategia di differenziazione già dalla sua fondazione (2017). Questo DB garantisce la monetizzazione ai suoi clienti tramite token di criptovaluta (DAT token) in cambio della condivisione dei loro dati. Gli utenti possono quindi vendere questi token su mercati di criptovalute o utilizzarli all'interno dell'ecosistema Datum; tuttavia, non ha ancora raggiunto una quantità sufficiente per poter competere attivamente contro i grandi CSP e gli altri DB.

Da questo modello, inoltre, si può comprendere perché molti partner dei DB siano proprio i CSP (Meta, X, ...); competere sull'appropriazione dei dati generati dai consumatori sui "social" non consentirebbe ai DB di ottenere sufficienti ricavi per sostenere la competizione.

1.8 RISCHI LEGATI ALLA FIGURA DEI DB

È evidente come alcune informazioni personali, cui i DB hanno accesso, se non opportunamente utilizzate, presentano intrinsecamente dei rischi per i consumatori stessi.

Per comprendere i suddetti rischi, vengono analizzate, di seguito, alcune pratiche di Oracle e Acxiom:

- **Oracle** negli ultimi 10 anni ha acquisito diverse aziende di dati, tra cui Datalogix, AddThis, Crosswise e BlueKai, piattaforme di gestione dei dati e mercati di dati. Datalogix aggrega dati su miliardi di transazioni di acquisto in 50 catene di supermercati e 1.500 grandi rivenditori, il servizio di social bookmarking AddThis traccia 900 milioni di utenti su 15 milioni di siti web, oltre a 1 miliardo di utenti mobili. Crosswise raccoglie i dati di attività su miliardi di dispositivi e identifica quali PC, telefoni, tablet e TV sono utilizzati da un singolo consumatore. Inoltre, Oracle aggrega e analizza 700 milioni di messaggi al giorno provenienti da social media, forum di discussione, blog, siti di recensioni dei consumatori e piattaforme video. Complessivamente, il marketplace di dati di Oracle fornisce più di 30.000 attributi di dati su due miliardi di profili di consumatori. Oracle classifica i consumatori in migliaia di categorie, che includono caratteristiche come età, genere, stato civile, composizione familiare, istruzione, occupazione, opinioni politiche, utilizzo dei media, reddito, prestiti, patrimonio e acquisti, ma anche dati sulle ricerche online. Ad esempio, Oracle tiene traccia di dati relativi all'etica personale e alla salute: tiene traccia di chi ha cercato argomenti come aborto, legalizzazione delle droghe, basi militari, proteste, insufficienza cardiaca o strutture mediche (caso Kochava). Insieme ai suoi partner di dati, l'azienda fornisce oltre 50.000 diverse categorie che possono essere assegnate ai consumatori. All'interno del suo data-cloud, Oracle registra "cosa fanno i consumatori", "cosa dicono i consumatori" e "cosa comprano i consumatori" per permettere alle aziende di trovare e indirizzare le persone su dispositivi e piattaforme, personalizzare le interazioni e misurare il comportamento dei consumatori. Le aziende possono importare i propri dati sui

clienti, visitatori del sito web e utenti delle app nel cloud Oracle, combinarli con dati e partner, e poi trasferirli e utilizzarli su centinaia di altri servizi di marketing e pubblicità. Inoltre, le aziende possono anche vendere i loro dati sui consumatori nel marketplace di Oracle.

Per identificare, collegare e abbinare i profili degli utenti tra diverse aziende, piattaforme, dispositivi e contesti in tempo reale, Oracle utilizza l'"Identity Graph" o "ID Graph". Questa tecnologia permette di unificare le identità indirizzabili su tutti i dispositivi. Essa "unisce tutte le interazioni su vari canali per creare un unico profilo consumatore indirizzabile", collegando diversi tipi di identificatori riferiti a individui, inclusi indirizzi postali, indirizzi e-mail, account utente per servizi online, ID mobili e ID di cookie. Quando i fornitori di dati o i clienti vogliono collegare i loro dati online, offline o mobili al cloud dei dati di Oracle, possono inviare "chiavi di abbinamento" che identificano i loro utenti "sia negli spazi online che offline". Oracle poi sincronizzerà queste chiavi di abbinamento con la sua rete di ID utente e ID statistici al suo cloud, che viene utilizzato per gestire ID e attributi utente. Per creare queste chiavi di abbinamento, i fornitori di dati e i clienti di Oracle devono convertire indirizzi e-mail, numeri di telefono, indirizzi postali, indirizzi IP e ID di dispositivi mobili in codici pseudonimi. Questi codici sono ancora riferimenti unici a determinati indirizzi e-mail, numeri di telefono e altri identificatori.

• **Acxiom** gestisce 15.000 database di clienti e 2,5 miliardi di relazioni con i clienti. Per quanto riguarda i dati dei consumatori, Acxiom sostiene di fornire accesso fino a 5.000 elementi di dati su 700 milioni di persone in tutto il mondo provenienti da "migliaia di fonti" in molti paesi. Acxiom assegna a ogni persona nel suo database un identificatore univoco a livello globale, proprio come Oracle. Questi ID personali all'interno del sistema "Abilitec Link" dell'azienda sono riconducibili a nome, indirizzo postale, numero di telefono, indirizzi e-mail e altri identificatori, e si basano su dati anagrafici e residenziali che risalgono a decenni fa. Acxiom raccoglie da lungo tempo dati da registri pubblici e da altre aziende, tra cui liste di abbonati a riviste, sondaggi e questionari dei consumatori, registrazioni di acquisti e garanzie.

Quando viene richiesta una persona specifica da un cliente, Acxiom fornisce, ad esempio, una tra 13 professioni religiose, tra cui "cattolico", "ebreo", "musulmano" e uno tra i quasi 200 "codici etnici". Complessivamente, l'azienda fornisce centinaia di attributi e punteggi sui consumatori; questi includono informazioni su istruzione, occupazione, figli, opinioni politiche, interessi, attività, acquisti, uso dei media, proprietà e veicoli posseduti, polizze bancarie e assicurative, reddito, prestiti, interessi sanitari e stabilità economica, nonché punteggi come "probabilità che qualcuno in una famiglia cambi carriera o lavoro". Dal 2014, con l'acquisizione della società di dati online LiveRamp, Acxiom ha compiuto sforzi significativi per collegare il suo database all'odierno universo di tracciamento e profilazione digitale. Grazie a tecnologie come l'"Identity Graph" e l'"IdentityLink" di LiveRamp, l'azienda afferma di essere in grado di collegare i dati offline e online a un unico

identificatore. Per raggiungere questo obiettivo, l'azienda abbina i profili dei consumatori agli identificatori online, come cookie e ID dei dispositivi. I clienti possono caricare i propri dati sui consumatori su LiveRamp, combinarli con i dati di oltre 100 fornitori di dati terzi, tra cui anche altri DB, come Equifax, Experian, TransUnion ed Epsilon, e poi utilizzarli su più di 500 piattaforme di tecnologia di marketing. Possono usare questi dati per trovare e indirizzare persone con caratteristiche specifiche, riconoscere e tracciare i consumatori su dispositivi e piattaforme, profilare e categorizzare, personalizzare i contenuti e misurare il loro comportamento.

I clienti potrebbero riconoscere un visitatore del sito web e fornire un'offerta personalizzata basata su dati di profilo estesi, senza richiedere all'utente di effettuare il login sul sito web. (Christl, 2017)

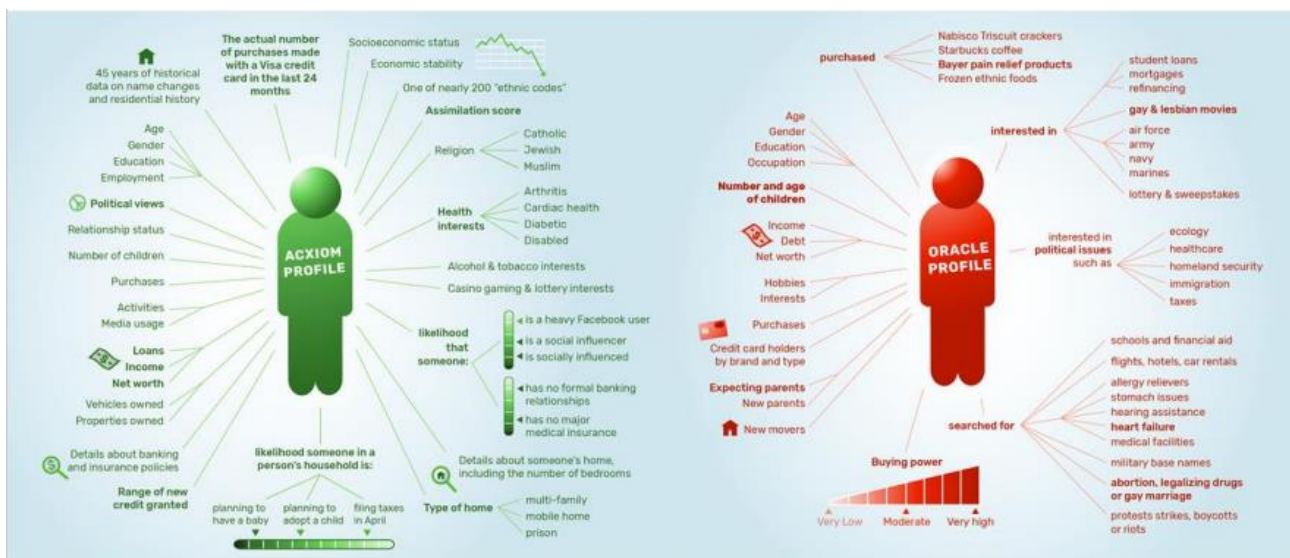


FIGURA 18: Acxiom e Oracle- tipologia di consumatori (Christl, 2017)

Da quando i DB hanno iniziato ad assumere un ruolo significativo per la concorrenza tra imprese, alcuni di essi hanno compiuto attività sui dati estratti che hanno allarmato le autorità e, ultimamente, anche tra i consumatori, vi è la percezione che le proprie informazioni siano facilmente accessibili a chiunque sia disposto a pagarle, forse anche a prezzi irrisori. Di seguito sono riportati esempi di news degli ultimi anni dal mercato statunitense dei DB:

- **LexisNexis** avrebbe raccolto oltre 78 miliardi di punti dati da 10.000 fonti pubbliche e private in 442 categorie di stili di vita sui rischi per la salute individuale e sui successivi costi medici. Sebbene l'azienda abbia dichiarato nel 2018 che la sua analisi non aveva ancora influenzato i costi assicurativi individuali, non vi è alcuna garanzia che questo non accadrà in futuro.
- A partire dal 2014, una sussidiaria di **Equifax** avrebbe aggregato e venduto l'accesso alle informazioni sui cedolini paga dei dipendenti per circa il 38 per cento della forza lavoro statunitense.

- **Datalogix** compila e vende la cronologia degli acquisti personali da oltre 1.400 programmi di fedeltà dei negozi.
- **SafeGraph, Ventel, X-Mode e Babel Street** acquistano geolocalizzazione e altre informazioni personali da app per smartphone. Negli ultimi anni, agenzie statunitensi tra cui l'FBI, il dipartimento della sicurezza interna degli Stati Uniti d'America (DHS) e l'Agenzia di Intelligence della Difesa (DIA) hanno acquistato informazioni di geolocalizzazione dagli smartphone statunitensi da data broker.
- **Clearview AI** avrebbe esaminato miliardi di immagini da siti web pubblicamente disponibili, inclusi i social media. Utilizzando tecnologie di riconoscimento facciale. Concede l'accesso a questo vasto database a oltre 3.000 entità federali e statali, tra cui la Central Intelligence Agency (CIA) e l'FBI.
- **Giant Oak, Palantir e Barbaricum** monitorano piattaforme di social media in aggregato, consentendo ai clienti di cercare miliardi di post generati dagli utenti per parole chiave o immagini. Il DHS avrebbe appaltato fornitori privati per tracciare i post di social media open source al fine di identificare individui che potrebbero essere associati a violazioni di visti, rischi per la sicurezza pubblica o minacce alla sicurezza nazionale.
- **Vigilant Solutions e Thomson Reuters** hanno monitorato e memorizzato la storia delle immagini e delle posizioni di miliardi di targhe in parcheggi all'aperto, autostrade e incroci, riuscendo a tracciare le auto degli individui per migliaia di agenzie di legge statunitensi.
- **Fog Data Science** ha pubblicizzato la raccolta di miliardi di segnali dati da app mobili, rivelando storie di posizioni "quasi in tempo reale" di oltre 250 milioni di dispositivi statunitensi risalenti al 2017. Ha venduto l'accesso a queste informazioni ad agenzie di polizia federali, statali e locali per meno di \$10.000 all'anno. (Center for Strategic and International Studies (CSIS), 2023)

SEZIONE 2 – POLICY

2.1 IL DATA ACT

Il documento di riferimento per il mercato dei dati è il Data Act: "Proposal for a REGULATION OF THE EUROPEAN PARLIAMENT AND OF THE COUNCIL on harmonised rules on fair access to and use of data (Data Act)". È stato pubblicato il 23 febbraio 2022 ed ha subito aggiornamenti nel corso del biennio 2022-2023; attualmente si attende un altro aggiornamento previsto per la fine del 2024. È finalizzato a stabilire regole armonizzate per garantire un accesso equo e trasparente ai dati per imprese, cittadini e istituzioni pubbliche, promuovendo l'innovazione e la crescita economica nell'Unione Europea. Tale documento mira a standardizzare e ottimizzare la gestione dei dati, sia nel settore pubblico che privato. Attraverso l'implementazione del Data Act, le autorità

governative sono tenute a seguire linee guida specifiche per la gestione e la divulgazione dei dati, garantendo una maggiore trasparenza nell'operato del governo e facilitando l'accesso ai dati da parte dei cittadini, delle imprese e degli altri enti pubblici. Ciò per favorire l'innovazione, l'analisi dei dati e la creazione di servizi pubblici più efficienti e orientati alle esigenze della società.

Il Data Act stabilisce standard comuni per la raccolta, la conservazione e la pubblicazione dei dati, promuovendo la disponibilità di dati in formati aperti e facilmente accessibili; appare, dunque, evidente come sia correlato all'operato dei DB. Tra gli obiettivi principali del Data Act vi è quello di creare un quadro normativo e un mercato chiari sulla gestione dei dati, garantendo al contempo un elevato livello di protezione dei dati personali e dei diritti fondamentali dei cittadini, facilitando così la libera circolazione dei dati all'interno dell'UE e promuovendo la cooperazione tra Stati membri e settori economici. È un atto incentrato sugli aspetti economici, commerciali del mercato dei dati che è perfettamente integrato con le altre normative europee sulla gestione dei dati.

- **Generazione dei dati**

Il Data Act sottolinea come la generazione dei dati sia un processo che coinvolge almeno due attori principali: il progettista o il fabbricante di un prodotto connesso e l'utente finale di tale prodotto o del servizio correlato. Questo implica che entrambi i soggetti giocano un ruolo cruciale nella creazione dei dati. Da un lato, i progettisti e i fabbricanti, che spesso sono anche fornitori di servizi correlati, sviluppano i dispositivi e le piattaforme che raccolgono i dati. Dall'altro lato, gli utenti interagiscono con questi prodotti e servizi, generando dati attraverso il loro utilizzo. Questa interazione solleva importanti questioni di equità nell'economia digitale. I dati registrati dai prodotti connessi e dai servizi correlati sono fondamentali per una vasta gamma di servizi post-vendita, ausiliari e di altro tipo. Pertanto, l'accesso a questi dati è cruciale non solo per i produttori originali, ma anche per altre imprese che potrebbero offrire servizi aggiuntivi, migliorare l'esperienza del cliente o innovare ulteriormente.

- **Protezione dei dati e Privacy**

Il diritto fondamentale alla protezione dei dati personali è garantito da regolamenti specifici dell'Unione Europea, in particolare dai regolamenti (UE) 2016/679 (GDPR) e (UE) 2018/1725. Il GDPR (Regolamento generale sulla protezione dei dati) è la principale normativa che stabilisce le condizioni per il trattamento dei dati personali all'interno dell'UE, assicurando che i dati siano trattati in modo legale, equo e trasparente.

Inoltre, la direttiva 2002/58/CE (Direttiva e Privacy) protegge la vita privata e la riservatezza delle comunicazioni elettroniche, stabilendo condizioni precise per l'archiviazione dei dati sia personali che non nei dispositivi terminali e per l'accesso a questi dati. Queste normative costituiscono la base per un trattamento dei dati sostenibile e responsabile, anche quando i set di dati includono una combinazione di dati personali e non personali.

Il Data Act integra queste normative esistenti, senza pregiudicarle; non modifica né limita i diritti alla protezione dei dati personali e alla riservatezza delle comunicazioni già stabiliti dalle normative dell'UE. Qualsiasi trattamento di dati personali effettuato ai sensi del Data Act deve rispettare il quadro normativo esistente dell'UE sulla protezione dei dati.

- **Minimizzazione e “privacy by design”**

Il Data Act sottolinea l'importanza di due principi fondamentali nella gestione dei dati: la minimizzazione dei dati e la protezione dei dati fin dalla fase di progettazione. Questi principi diventano critici quando il trattamento dei dati implica rischi significativi per i diritti fondamentali delle persone. La minimizzazione dei dati implica la raccolta e il trattamento solo delle informazioni strettamente necessarie per il loro scopo specifico, riducendo così il rischio di utilizzo improprio o di violazione della privacy. La protezione dei dati fin dalla fase di progettazione, o "privacy by design", richiede che le misure di sicurezza e di tutela dei diritti fondamentali siano integrate nel design stesso del sistema o del processo di trattamento dei dati, anziché essere aggiunte in un secondo momento. Questo approccio preventivo mira a prevenire potenziali violazioni della privacy e altri rischi per i diritti delle persone sin dalle fasi iniziali di sviluppo. Tutte le parti coinvolte nella condivisione dei dati devono adottare misure tecniche e organizzative adeguate per garantire la protezione dei diritti fondamentali. Queste misure non si limitano alla pseudonimizzazione e alla cifratura (che protegge i dati attraverso un algoritmo crittografico), ma includono anche l'adozione di tecnologie avanzate che consentono di analizzare i dati o copiarli in maniera non necessaria.

2.2 LA REGOLAZIONE DEI DB

I dibattiti giuridici e politici riguardanti i data broker – inizialmente sull'intermediazione di punteggi di credito – risalgono agli anni '70. Per decenni, i data broker hanno operato in un mercato con poca o nessuna regolamentazione, dove le transazioni tra aziende e governi venivano condotte senza restrizioni o controllo pubblico. In presenza di una regolazione minima, molte pratiche erano possibili senza subire alcun controllo a riguardo: acquisire, acquistare, vendere, condividere informazioni, comprese quelle estratte o dedotte statisticamente. La richiesta di una regolazione più forte ha avuto origine quando è nata la preoccupazione di “incursioni” nella privacy da parte del settore dei data broker. Con il passare del tempo, visti i vari scandali avvenuti negli ultimi dieci anni (in particolare lo scandalo Cambridge Analytica), nonché la ricerca critica sui rischi che i big data comportano per i diritti umani, l'attenzione delle autorità governative, antitrust e lentamente anche del pubblico di tutto il mondo si sta rivolgendo anche alla figura dei DB. Nonostante tali discussioni, si ha il dubbio che i DB rimangano per lo più non regolamentati, ad esempio: il California Consumer Privacy Act (CCPA) consente ai residenti di ottenere copie delle informazioni trovate su di loro dai data broker. Il disegno di legge consente inoltre ai

consumatori della California di richiedere la cancellazione delle informazioni e scegliere di non vendere i propri dati; oppure, negli Stati Uniti, dove risiedono la maggior parte dei DB più influenti a livello mondiale, molti di essi “non riconoscono praticamente alcun diritto” ai consumatori nelle loro politiche, mentre alcuni includono anche una clausola che riserva il diritto dell'azienda di modificare i propri standard sui dati in qualsiasi momento. Negli USA sono state discusse numerose proposte legislative. La “legge sulla trasparenza delle informazioni e sul controllo dei dati personali” avrebbe richiesto ai DB di ottenere il consenso per raccogliere dati sensibili e sottoporsi a un controllo annuale sulla privacy. La “legge sulla responsabilità e la fiducia dei dati” avrebbe stabilito standard di sicurezza e richiesto controlli post-violazione degli intermediari di dati, oltre a proibire la raccolta di informazioni con false pretese. La “legge sull'elenco dei DB” del 2019 avrebbe richiesto ai data broker di iscriversi a un registro nazionale supervisionato dalla FTC e di mantenere un programma completo di sicurezza delle informazioni per proteggere i dati dei consumatori da violazioni della sicurezza e altre divulgazioni involontarie o improprie. Similmente, lo Stato di New York sta valutando un proprio disegno di legge per obbligare i data broker a registrarsi nel proprio Stato. Il disegno di legge darà inoltre istruzioni al procuratore generale dello stato di mantenere uno specifico sito web che elenchi questi intermediari.

In Europa, la regolazione che riguarda i DB più da vicino è il GDPR, presentato nel prossimo paragrafo.

I dati sono multidimensionali, fluidi e hanno proprietà insolite. Le loro caratteristiche ne sfidano la classificazione e, di conseguenza, la regolamentazione e la protezione. Sono beni non-rivali: possono essere utilizzati da più entità contemporaneamente senza esaurirsi, il che li rende molto diversi da altre risorse materiali; e sono escludibili: possono essere controllati e l'accesso ad essi può essere limitato da chi ha il potere di farlo; tuttavia, questo diritto raramente viene esercitato dai soggetti dei dati. Pertanto, i dati sono spesso frammentati o copiati, con diverse parti memorizzate in luoghi differenti, costantemente smontate e rimontate, e spostate tra server. Di conseguenza, gli utenti non hanno controllo su dove i loro dati sono memorizzati e spesso non sanno dove si trovano né quale giurisdizione li governi. Fondamentalmente, i dati, se non sufficientemente regolati, sfidano il ruolo del territorio come epistemologia spaziale di base nel diritto internazionale (Reviglio, 2022).

2.3 IL GDPR

Il General Data Protection Regulation (GDPR) è il regolamento che tocca maggiormente, anche se non direttamente, la figura dei DB ed è, dunque, quello di maggiore interesse per il presente elaborato. Nel GDPR si definisce “«trattamento»: qualsiasi operazione o insieme di operazioni [...] applicate a dati personali o insiemi di dati personali”; si definisce: “«titolare del trattamento»: la persona fisica o giuridica, l'autorità pubblica, il servizio o altro

organismo che [...] determina le finalità e i mezzi del trattamento di dati personali”; e si definisce “«consenso dell'interessato»: qualsiasi manifestazione di volontà libera, specifica, informata e inequivocabile dell'interessato, con la quale lo stesso manifesta il proprio assenso [...] che i dati personali che lo riguardano siano oggetto di trattamento”

- **Oggetto e finalità del GDPR**

Il GDPR stabilisce norme riguardanti la protezione delle persone fisiche con riferimento al trattamento dei dati personali, nonché norme relative alla libera circolazione di tali dati; mira a proteggere i diritti e le libertà fondamentali degli individui, in particolare il diritto alla protezione dei dati personali. Un aspetto cruciale del GDPR è che la libera circolazione dei dati personali all'interno dell'Unione Europea non può essere limitata né vietata per motivi legati alla protezione delle persone fisiche con riguardo al trattamento dei dati personali.

Il GDPR è stato introdotto con l'obiettivo di armonizzare le leggi sulla protezione dei dati all'interno dell'UE, garantendo così un livello coerente di protezione per gli individui in tutti gli Stati membri. Inoltre, il regolamento cerca di facilitare la circolazione dei dati personali all'interno dell'Unione, promuovendo la fiducia e la sicurezza nelle transazioni digitali e nei servizi transfrontalieri.

- **Ambito di applicazione materiale**

L'ambito di applicazione materiale del GDPR comprende qualsiasi operazione o insieme di operazioni effettuate su dati personali, sia con mezzi automatizzati sia con mezzi manuali, purché i dati facciano parte di un archivio strutturato. Questo approccio ampio e inclusivo è progettato per garantire che le norme di protezione dei dati siano applicabili in vari contesti e settori, adattandosi alle diverse modalità con cui le informazioni personali possono essere gestite.

- **Ambito di applicazione territoriale**

L'ambito di applicazione territoriale del GDPR è progettato per garantire che la protezione dei dati personali sia mantenuta costantemente, indipendentemente dalla posizione geografica del trattamento dei dati. Questo è particolarmente rilevante nel settore dei DB, settore in cui i player più importanti hanno sedi operative nell'UE, ma sedi centrali negli Stati Uniti. In generale, qualsiasi organizzazione che ha uno stabilimento nell'UE e/o tratta dati personali all'interno dell'unione deve conformarsi alle disposizioni del GDPR, anche se il trattamento stesso avviene in un paese terzo.

- **Principi applicabili al trattamento di dati personali**

Il GDPR assicura che i dati siano trattati in modo lecito, corretto e trasparente nei confronti dell'interessato, garantendo così la liceità, correttezza e trasparenza del trattamento. Uno dei principi chiave del GDPR (come per il Data Act citato precedentemente) "minimizzazione

dei dati", che richiede che i dati personali siano adeguati, pertinenti e limitati alle finalità per le quali sono trattati. Questo principio mira a evitare l'accumulo eccessivo di dati e a garantire che solo le informazioni strettamente necessarie siano raccolte e utilizzate. Inoltre, i dati personali devono essere esatti e, se necessario, aggiornati. Questo principio garantisce che le informazioni trattate siano accurate e riflettano la situazione attuale dell'interessato. Infine, i dati personali devono essere trattati in maniera da garantire un'adeguata sicurezza, inclusa la protezione mediante misure tecniche e organizzative adeguate.

▪ **Condizioni per il consenso**

Il GDPR pone un forte accento sull'importanza del consenso dell'interessato come base giuridica per il trattamento dei dati personali. Il titolare del trattamento deve essere in grado di dimostrare che l'interessato lo abbia effettivamente prestato in relazione al trattamento dei suoi dati personali. Tale principio garantisce che il consenso sia documentato e verificabile, fornendo trasparenza e responsabilità nel trattamento dei dati. Un altro aspetto fondamentale è il diritto dell'interessato di revocarlo in qualsiasi momento. Tale revoca non pregiudica la liceità del trattamento basata sul consenso prima della revoca. È essenziale che l'interessato sia informato di questo diritto prima di prestarlo e che il processo di revoca sia semplice quanto quello di accettazione.

▪ **Trattamento di Categorie Particolari di Dati Personali**

Il GDPR vieta il trattamento di categorie particolari di dati personali che "rivelino l'origine razziale o etnica, le opinioni politiche, le convinzioni religiose o filosofiche, l'appartenenza sindacale, nonché dati genetici, dati biometrici intesi a identificare in modo univoco una persona fisica, dati relativi alla salute o alla vita sessuale o all'orientamento sessuale della persona".

Esistono, tuttavia, alcune eccezioni a questo divieto, che permettono il trattamento di tali dati in specifiche circostanze, ad esempio:

- a. il trattamento è consentito se l'interessato ha prestato il proprio consenso esplicito al trattamento di tali dati personali per una o più finalità specifiche. Ciò garantisce che il soggetto sia pienamente consapevole e d'accordo con l'uso dei propri dati sensibili.
- b. il trattamento è necessario per assolvere gli obblighi ed esercitare i diritti specifici del titolare del trattamento o dell'interessato in materia di diritto del lavoro, sicurezza sociale e protezione sociale.
- c. il trattamento è necessario per tutelare un interesse vitale dell'interessato, soprattutto quando quest'ultimo non è in grado di dare il proprio consenso.
- d. il trattamento è effettuato da una fondazione, associazione o altro organismo senza scopo di lucro che persegue finalità politiche, filosofiche, religiose o sindacali, nell'ambito delle sue legittime attività e con adeguate garanzie.

- e. il trattamento riguarda dati personali resi manifestamente pubblici dall'interessato, implicando che l'interessato ha volontariamente divulgato tali informazioni.
- f. il trattamento è necessario per accertare, esercitare o difendere un diritto in sede giudiziaria, assicurando che le informazioni possano essere utilizzate per proteggere i diritti legali.

- **Principali diritti dell'interessato**

L'articolo 15 (diritto di accesso dell'interessato) garantisce all'interessato "il diritto di ottenere dal titolare del trattamento la conferma che sia o meno in corso un trattamento di dati personali che lo riguardano e in tal caso, di ottenere l'accesso ai dati personali". L'interessato può richiedere, tra le altre, le seguenti informazioni: le finalità del trattamento, le categorie di dati personali in questione, i destinatari a cui i dati personali son già stati o verranno comunicati, ove possibile, il periodo di conservazione dei dati personali previsto oppure, se non fosse possibile, i criteri utilizzati per determinare tale periodo.

L'articolo 16 (diritto di rettifica) garantisce "il diritto di ottenere dal titolare del trattamento la rettifica dei dati personali inesatti che lo riguardano senza ingiustificato ritardo".

L'articolo 17 (diritto alla cancellazione) garantisce il diritto di ottenere dal titolare del trattamento la cancellazione dei dati personali che lo riguardano se vi è la presenza di almeno uno dei seguenti motivi: i dati personali non sono più necessari rispetto alle finalità per le quali sono stati raccolti; l'interessato revoca il consenso; l'interessato si oppone al trattamento; i dati personali sono stati trattati illecitamente;

L'articolo 21 (diritto di opposizione) garantisce all'interessato il diritto di opporsi in qualsiasi momento al trattamento dei dati personali che lo riguardano. Il titolare del trattamento si astiene dal trattare ulteriormente i dati personali salvo che egli dimostri l'esistenza di motivi legittimi cogenti per procedere al trattamento che prevalgono sugli interessi, sui diritti e sulle libertà dell'interessato oppure per l'accertamento, l'esercizio o la difesa di un diritto in sede giudiziaria

- **Altri articoli del GDPR sono:**

L'articolo 32 afferma che il titolare del trattamento deve mettere in atto misure tecniche e organizzative adeguate per garantire un livello di sicurezza adeguato al rischio. Tali misure devono tenere conto dello stato dell'arte, dei costi di attuazione, della natura, dell'oggetto, del contesto e delle finalità del trattamento. Le misure di sicurezza devono essere proporzionate ai rischi che il trattamento dei dati comporta e possono includere le seguenti azioni:

a) Pseudonimizzazione e cifratura dei dati personali

- b) Riservatezza, integrità, disponibilità e resilienza dei sistemi e dei servizi di trattamento: è essenziale assicurare su base permanente che i sistemi e i servizi di trattamento mantengano la confidenzialità, l'integrità e la disponibilità dei dati.
- c) Ripristino tempestivo dei dati personali: le organizzazioni devono essere in grado di ripristinare rapidamente la disponibilità e l'accesso ai dati personali in caso di incidenti fisici o tecnici, garantendone la continuità operativa.
- d) Test, verifica e valutazione regolari

L'articolo 34 (comunicazione di una violazione dei dati personali all'interessato) parla di violazione dei dati personali; una violazione dei dati personali si verifica quando vi è un incidente che compromette la sicurezza dei dati in questione. Ciò può includere situazioni come la perdita, l'accesso non autorizzato, la divulgazione o la distruzione di dati personali che possono avere conseguenze dannose per le persone fisiche interessate. Questi danni possono essere di natura fisica, materiale o immateriale, come la perdita di controllo sui propri dati, restrizioni ai propri diritti, discriminazione, furto di identità, perdite finanziarie, danni alla reputazione, o altri significativi danni economici o sociali.

L'articolo afferma che se vi è presenza di una violazione dei dati personali, e tale violazione è suscettibile di presentare un rischio elevato per i diritti e le libertà delle persone fisiche, il titolare del trattamento deve comunicare la violazione all'interessato senza ingiustificato ritardo. In determinate situazioni, la notifica all'interessato potrebbe non essere necessaria se almeno una delle seguenti condizioni è soddisfatta:

- a) Il responsabile del trattamento ha implementato adeguate misure tecniche e organizzative di protezione, che includono la pseudonimizzazione e la cifratura, per rendere i dati personali incomprensibili a individui non autorizzati che potrebbero accedervi;
- b) Il responsabile del trattamento ha preso successivamente misure volte a mitigare un rischio elevato per i diritti e le libertà degli interessati;
- c) La notifica agli interessati richiederebbe sforzi sproporzionati. In tal caso, può essere effettuata una comunicazione pubblica o un'azione simile che garantisca agli interessati un'informazione efficace e comparabile.

2.4 DB E GDPR

Ultimamente, come accennato nella sezione 1, i DB sono entrati sempre di più nel mirino delle autorità per comprendere se il loro modello di business fosse compatibile e correttamente regolamentato rispetto alle legislazioni sulla digitalizzazione (ad esempio il GDPR) o se occorre crearne una ad hoc. Dal punto di vista giuridico, i data broker sono

individuati come soggetti che traggono un vantaggio economico, sia diretto che indiretto, dalle transazioni riguardanti dati personali e non, senza coinvolgere direttamente l'interessato. Questa definizione ha rilevanza giuridica poiché determina l'applicabilità della legislazione, come il GDPR, e il conseguente equilibrio degli interessi tra il soggetto dei dati e l'intermediario (Ruscheimer, 2023). Alla luce della definizione giuridica di DB, bisogna tenere in considerazione che più enti potrebbero rientrarvi, come a esempio molti CSP.

Prendendo in considerazione il mercato europeo, anche se considerazioni analoghe possono essere fatte per gli altri continenti in cui operano, i DB potrebbero rappresentare una minaccia per la protezione dei dati e la privacy degli individui. L'avversione a tali modelli di business è in crescita poiché sono percepiti come una conseguenza di un settore, quello dei dati, caratterizzato da asimmetrie informative e di eccessivo potere economico, probabilmente non sufficientemente regolato.

Le autorità tedesche per la protezione dei dati hanno recentemente suggerito di vietare la vendita dei dati commerciali a causa delle disposizioni del GDPR; quelle francese e danese, similmente, hanno intrapreso attività di vigilanza proprio sui modelli dei DB.

I DB operano solo nei casi in cui possono sistematicamente impedire ai soggetti dei dati di esercitare i diritti precedentemente citati, principalmente attraverso una mancanza di trasparenza (che, ammesso sia presente, potrebbe non essere sufficiente) o chiarezza; l'interessato, infatti, di solito non è a conoscenza della vendita dei propri dati, non è coinvolto nel contratto di vendita e non ha la possibilità di negoziare condizioni che ne vietano l'uso secondario.

Le autorità antitrust ritengono che il commercio di indirizzi commerciali senza il consenso dei soggetti interessati non sia consentito ai sensi del GDPR. Inoltre, né il soggetto dei dati né l'intermediario che li vende sanno a chi potrebbe essere venduto l'indirizzo in futuro. Pertanto, questo scambio probabilmente si dovrebbe scontrare con una barriera legale. Se questa valutazione è corretta, ciò avrà un impatto sull'intera applicazione del GDPR riguardante il modello di business dei data broker che potrebbe, appunto, non essere conforme al GDPR, a causa dei problemi legati all'ottenimento del consenso informato negli ambienti digitali e della necessità di bilanciarlo con l'interesse legittimo come base per il trattamento lecito dei dati personali.

Il modello di business degli intermediari di dati non è una "tecnologia digitale innovativa" (prova ne è che le prime società di data-brokeraggio sono state fondate negli anni 50 dell'800), ma piuttosto un metodo percepito come meno invasivo, e a volte non percepito affatto, rispetto alla pubblicità postale analogica e alle chiamate telefoniche indesiderate, ovvero la pubblicità personalizzata su internet.

Tenendo presente la natura dei dati e la conseguente difficoltà nel controllarne la provenienza dopo varie modifiche, aggregazioni, partizionamenti, statistiche e inferenze effettuate, come già esplicitato per il GDPR, le normative si applicano secondo il principio

lex loci solutionis, estendendo la loro giurisdizione alle operazioni dei data broker che hanno sede al di fuori dell'UE e che si rivolgono ai cittadini europei. Di conseguenza, anche DB statunitensi, come Acxiom, sono soggetti alla regolazione europea quando operano in Europa.

▪ Il modello di business dei DB

I data broker operano nel campo dell'uso dei dati secondari, il loro modello di business ha inizio dopo che gli operatori dei dati, ad esempio i content service providers, hanno ricevuto dati dai soggetti stessi tramite le metodologie espresse nella sezione 1. Essi stipulano un contratto con i loro acquirenti e lo eseguono vendendo o noleggiando i dati. Le copie dei dati, come altre materie prime, spesso vengono vendute a un prezzo molto basso, a diversi acquirenti. Questi acquirenti estraggono ulteriori informazioni dai set di dati tramite data mining, machine learning e analisi predittiva; di conseguenza, i data broker non hanno interessi specifici, nella fase di raccolta e di vendita, correlati all'uso per cui verranno utilizzati i dati; il loro modello di business inizia e termina con la vendita o noleggio dei dati. Tuttavia, benché l'utilizzo finale non sia di interesse dei DB, ciò non significa che tali intermediari non debbano essere considerati stakeholder nella creazione di fallimenti di mercato o problematiche collegate ai dati da loro utilizzati: anche la sola commercializzazione dei dati contribuisce e peggiora questi problemi.

Come descritto in precedenza, i data broker vendono dati ai clienti segmentando in base all'etnia, al reddito, allo stato di salute, all'orientamento sessuale e ad altre informazioni sensibili di una persona. Da un punto di vista legale, ciò solleva due questioni chiave:

1. la violazione della legislazione sulla privacy e sulla protezione dei dati e,
2. la discriminazione non trasparente.

Anche la scienza politica sostiene che i DB minano la stabilità geopolitica (caso TikTok) e la fiducia nei mercati dei dati. Trattare questa tipologia di dati come della merce (commodification) che può essere scambiata liberamente, crea problemi significativi; in quanto, essi sono correlati a tematiche di protezione della privacy sia a livello individuale che collettivo. Inoltre, i DB alimentano asimmetrie informative vendendo dati a banche, datori di lavoro, compagnie assicurative e governi, pratiche dalle quali i consumatori attualmente non sono sufficientemente protetti (tematiche correlate sono l'azzardo morale o la selezione avversa). Questo in parte perché il mercato dei DB non sembra essere sufficientemente chiaro e trasparente; ad esempio, non esiste, ancora, un elenco di tali aziende nell'UE, e per fare ciò si potrebbe prendere come riferimento la "legge sull'elenco dei DB" statunitense del 2019 citata in precedenza. Le leggi attuali non riescono a catturare tali pratiche poiché i data broker sono attivi a livello globale e profondamente integrati nell'ecosistema digitale (Ruscheimer, 2023).

Un altro fattore da tenere in considerazione il fatto che certi DB, anche tra i più influenti, hanno fornito sui dati che vendevano un tasso di accuratezza solo del 50%, con conseguenti derivazioni e profili falsi e ingiustificati. Questo è ancora più problematico nell'era dei big data, dove tutti i tipi di informazioni apparentemente innocue possono portare a inferenze sensibili sulle persone, specialmente in termini di accesso al finanziamento o alle cure mediche.

Il problema dell'inferenza che può essere tratta dai dati personali è meglio esemplificato da una delle peggiori violazioni dei dati nella storia, eseguita contro Equifax, uno data broker statunitense che è anche una delle tre agenzie di credito statunitensi.

Nel 2017, Equifax ha annunciato una violazione di dati che ha esposto i dati personali di 147 milioni di persone, inclusi quelli dei clienti nel Regno Unito e nell'Unione Europea. La società, per affrontare l'enorme problema, ha stipulato un accordo con la Federal Trade Commission, il Consumer Financial Protection Bureau e 50 stati e territori degli Stati Uniti. L'accordo prevedeva fino a 425 milioni di dollari per aiutare le persone colpite dalla violazione dei dati.

Vi sono ripercussioni ancora oggi, infatti, sul sito della FTC, si legge: "si può presentare una richiesta di risarcimento per le spese sostenute tra il 23 gennaio 2020 e il 22 gennaio 2024 a seguito di furto di identità o frode correlata alla violazione, come:

- perdite derivanti da addebiti non autorizzati sui conti
- commissioni pagate a professionisti, come commercialisti o avvocati, per aiutare i consumatori nel caso di furto di identità
- Altre spese sostenute durante il recupero dal furto di identità, come spese notarili, spese di spedizione dei documenti, spese postali, ed altre" (FTC, 2024).

Tutt'ora, mancano incentivi sufficienti per i data broker per crittografare e proteggere i dati su cui lavorano.

Un altro fattore che contribuisce alla carenza di trasparenza potrebbe essere il fatto che il lavoro dei DB è strettamente legato alla piattaforma o sito da cui estraggono i dati. Poche società posseggono i social media utilizzati dalla maggioranza degli individui del globo e tali aziende hanno ottimizzato il design delle loro interfacce in modo che gli utenti generino quanti più dati possibile, questi meccanismi non sono facilmente identificabili e i DB riescono a generare moltissimo valore da queste piattaforme.

Nel lavoro "Data Brokers and European Digital Legislation" (Ruscheimer, 2023) vengono analizzati i determinati concetti di interesse giuridico sulla figura dei DB:

- Consenso

A livello pratico, la base legale più rilevante per la liceità del trattamento dei dati negli ambienti digitali è il consenso, Art. 6 GDPR. Va notato che, su Internet, i requisiti del

consenso informato e volontario non vengono soddisfatti nella maggior parte dei casi. Inoltre, l'era digitale rende talvolta il consenso una "finzione". Gli utenti di Internet non possono tenere traccia dei numerosi attori e scopi di trattamento coinvolti, in particolare, i moduli di consenso user-friendly coprono una quantità di informazioni così grande per ogni sito web che prendere decisioni informate diventa sostanzialmente impossibile. Nel recente caso del transparency and consent framework (TCF), uno standard sviluppato dall'Interactive Advertising Bureau (IAB) Europe per aiutare gli editori, gli inserzionisti e altri operatori del settore della pubblicità online a ottenere il consenso degli utenti per la raccolta e l'elaborazione dei dati personali in conformità con il GDPR, l'autorità belga per la protezione dei dati ha osservato che il TCF rende difficile per gli utenti ottenere maggiori informazioni sull'identità di tutti i "controllori dei dati" a cui stanno dando consenso prima che tale consenso venga dato. In particolare, la DPA (Data Protection Authority) belga ha sostenuto che i numerosi destinatari del consenso richiederebbero agli utenti di spendere un tempo sproporzionato per leggere tutte le informazioni necessarie, il che significa che il consenso raramente potrebbe essere sufficientemente informato. Per quanto riguarda i data broker: i soggetti dei dati probabilmente non prevederanno le conseguenze del dare il consenso alla vendita dei propri dati individuali, specialmente in relazione ai servizi connessi, ad esempio, nell'Internet of things. Il consenso informato richiederebbe che il soggetto dei dati fosse informato in anticipo su chi vende i dati, per quali scopi, se quel compratore rivenderà i dati, ecc. Sebbene sia teoricamente possibile perseguire il modello di business dei data broker in conformità con il GDPR, ciò richiederebbe condizioni preliminari per il consenso informato da parte del soggetto dei dati a tutti gli scopi per i quali i dati verranno trattati. La vendita dei set di dati da parte dei data broker e le analisi predittive influenzano una enorme mole di persone per definizione, rendendo queste condizioni apparentemente impossibili da soddisfare.

- Marketing diretto

Per quanto riguarda il marketing diretto, il diritto di opposizione del soggetto dei dati è garantito dall'articolo 21 del GDPR come analizzato in precedenza; tuttavia, poiché i data broker non si occupano direttamente del marketing, ma piuttosto vendono informazioni su segmenti di consumatori, il diritto di opposizione del soggetto dei dati non può essere applicato a queste figure. Ciò significa che, non essendo i DB titolari del trattamento, sarà più difficile per il soggetto dei dati far valere il proprio diritto di opposizione su di loro.

- L'interesse commerciale legittimo dei data broker

La definizione di "interessi legittimi" è estesa in modo che possa coprire una vasta gamma di interessi, inclusi quelli legali, economici e idealistici. Gli interessi sono considerati legittimi quando rispettano le norme del diritto della protezione dei dati o del sistema legale in generale. Ciò implica che non basta che un'attività sia legalmente permessa, ma deve anche essere giustificata da altri principi e norme. Il trattamento dei dati, da parte dei DB, deve essere necessario affinché l'interesse sia considerato legittimo.

Vi è un dibattito su come il GDPR valuti gli interessi puramente commerciali: il GDPR e il diritto alla privacy non definiscono un diritto economico esclusivo dell'individuo a commercializzare i propri dati. Mentre è legittimo e spesso desiderabile dal punto di vista legale gestire modelli di business orientati al profitto, i modelli di business dei data broker (che vendono dati personali) sembrano essere in contrasto con il requisito di necessità.

- Necessità e legittimità

Fare affidamento sulla base legale dell'interesse legittimo per giustificare il trattamento dei dati a fini di monetizzazione è problematico. L'efficienza e l'urgenza da sole possono dimostrare un interesse legittimo, ma non possono soddisfare il secondo requisito, ossia che il trattamento dei dati sia necessario. In linea di massima, gli interessi del responsabile del trattamento dei dati, in questo caso il data broker, sono opposti agli interessi dell'interessato riguardanti i loro diritti fondamentali alla protezione dei dati.

Al contrario dell'interesse legittimo, il criterio di necessità è interpretato piuttosto ristrettamente. Il trattamento dei dati è necessario solo quando l'interesse legittimo non può essere raggiunto con altri mezzi; infatti, la giurisprudenza della Corte di Giustizia dell'Unione Europea è limitata a ciò che è assolutamente necessario.

I data broker possono perseguire interessi puramente commerciali con meno dati: dati temporanei, dati non collegati, ecc. Tutti questi fattori devono essere interpretati alla luce dei principi del diritto della protezione dei dati, che sottolinea il fatto che il commercio dei big data non è conforme al GDPR. Inoltre, la possibilità di trarre inferenze sui soggetti dei dati dall'insieme di dati, il livello di connettività e la durata della conservazione dei dati, influenzano il diritto alla protezione dei dati dell'individuo e devono essere presi in considerazione. Queste considerazioni indeboliscono la posizione dei DB e rafforzano quella degli interessati nel test di necessità.

Altri problemi legali legati alla figura dei DB sono: l'Art. 5 del GDPR che richiede che i dati siano accurati e aggiornati, il che non è necessariamente nell'interesse primario delle società di brokeraggio dati. Inoltre, i dati devono essere elaborati in modo comprensibile per il soggetto dei dati; tuttavia, è, di fatto, impossibile per la persona individuale comprendere come i dati scambiati dai DB vengano raccolti ed elaborati, quando si tratta di una mole così grande di informazioni. In questa ottica, il segreto che circonda l'aggregazione dei dati interferisce ulteriormente con i diritti dell'interessato. Nel sistema dei diritti individuali ai sensi del GDPR, ciò comporta la perdita del controllo dei propri dati da parte del soggetto dei dati.

In conclusione, il commercio dei dati per interessi puramente commerciali di solito non è conforme al GDPR. Il problema della difficoltà nell'attribuzione della applicazione amplifica questa questione, poiché i soggetti dei dati possono far valere i loro diritti contro le violazioni della protezione dei dati solo se ne sono consapevoli.

L'anonimizzazione e la trasparenza non sembrano essere la soluzione:

- l'anonimizzazione dei dati, spesso pubblicizzata come la soluzione al problema della privacy, non risolve il problema delle asimmetrie informative tra gli attori commerciali dell'industria dei dati e il soggetto dei dati. Al contrario, le promesse di dati anonimizzati portano gli utenti, che non sono consapevoli della loro applicazione collettiva, a un falso senso di sicurezza. Inoltre, spesso, dati ritenuti anonimi sono in realtà identificabili e riconducibili;
- anche il requisito di trasparenza non è adatto per limitare gli eccessi del commercio dei dati, poiché i DB non raggiungeranno mai una trasparenza che garantisca una scelta consapevole da parte dei soggetti. Maggiori informazioni sulle opzioni decisionali non portano linearmente a un maggiore controllo; infatti, la trasparenza da sola non è sufficiente, è necessaria la comprensibilità e la libera scelta di diverse opzioni. Il concetto di "auto-pericolo informativo" calcolato può applicarsi solo se sono disponibili informazioni complete, coerenti e comprensibili come base per la presa di decisioni.

La trasparenza completa e la divulgazione di tutti i fattori rilevanti da sole non raggiungono una decisione informata quando non sono accompagnate dalla comprensione. La complessità del trattamento dei dati nei servizi online rende questo illusorio. La quantità di informazioni necessarie non porterebbe a una comprensione reale, ma al "paradosso della trasparenza" in cui la chiarificazione richiesta per ottenere trasparenza richiede una quantità di informazioni che rende difficile per l'utente medio comprenderne la portata; come è noto, soltanto leggere le condizioni dei siti web che un utente visita richiederebbe una notevole quantità di tempo - calcolato per 201 ore in media all'anno (McDonald, et al., 2008). È infatti chiaro che il concetto di autogestione della privacy ha perso gran parte della sua efficacia.

Per quanto riguarda le nuove legislazioni europee, esse "non sembrano occuparsi delle attività dei DB":

Anche se la Direttiva Armonizzata (DA) mira a modellare l'economia europea dei dati, il suo campo di applicazione è limitato e probabilmente non influenzerà il modello di business dei data broker. La DA stabilisce regole uniformi per l'accesso ai dati generati dall'Internet of things e dai servizi correlati, a beneficio di consumatori, utenti commerciali e autorità pubbliche. Definisce anche le specifiche per la redazione dei contratti quando un titolare dei dati trasferisce dati a terzi per soddisfare richieste di accesso ai dati. Inoltre, la DA ha obiettivi diversi dalla protezione degli interessi legali messi in pericolo dai DB. Anche se ci sono soggetti interessati che traggono vantaggio dall'uso dei propri dati, l'obiettivo principale è rimuovere le barriere all'accesso.

Il Data Governance Act (DGA) mira a stabilire un quadro chiaro per l'uso dei dati nel settore pubblico. Questi dati possono includere segreti commerciali, dati personali o proprietà intellettuale. Tuttavia, il DGA non influisce sul modello di business dei data broker in

quanto disciplina principalmente le condizioni per l'accesso ai dati da parte delle autorità pubbliche e la condivisione volontaria dei dati. Al contrario, il modello di business dei data broker si basa sull'estrazione "aggressiva e non trasparente" dei dati da varie fonti, anziché sulla semplice condivisione dei dati. Inoltre, il DGA chiarisce esplicitamente che non mira a interferire con il modello di business dei data broker.

La European Data Spaces (EDS) è un'iniziativa promossa dalla Commissione Europea nell'ambito della strategia europea per i dati. L'obiettivo principale è creare un'infrastruttura comune che consenta la condivisione e l'utilizzo dei dati tra le diverse organizzazioni, settori e Stati membri dell'Unione Europea. La EDS mira a promuovere l'innovazione, l'efficienza e la competitività nel mercato unico digitale europeo, con l'obiettivo di rispettare gli standard di privacy e protezione dei dati; tuttavia, i DB potrebbero trarre a proprio vantaggio queste iniziative per ampliare i propri database con informazioni sensibili. Uno degli obiettivi è "rendere più semplice scambiare dati sanitari e accedervi a livello dell'UE", questo potrebbe permettere ai DB di ottenere informazioni sanitarie sui cittadini ancora più facilmente e potrebbe sfociare in casi paragonabili allo scandalo di Kochava negli USA.

Il paper di Ruschemeier si conclude affermando che, comunque, la nuova legislazione dell'UE rappresenta un passo nella direzione corretta; tuttavia, manca un quadro che regoli in modo efficace le asimmetrie di potere tra i responsabili del trattamento dei dati, come i DB, e gli interessati. Un primo passo potrebbe essere una definizione univoca e la registrazione dei data broker e, soprattutto, l'istituzione di regole di trasparenza, come richiesto dalla DSA per le grandi piattaforme online.

2.5 IL CAPITALISMO DI SORVEGLIANZA

L'espressione "capitalismo di sorveglianza", ideata da Shoshana Zuboff, sintetizza efficacemente due concetti: quello di un capitalismo nuovo, differente da quello industriale dei secoli scorsi, e quello di un sistema di potere emergente basato sul controllo del comportamento individuale.

Nei primi anni 2000, Google scopre il principio di questo nuovo capitalismo. L'idea di base è quella di "appropriarsi dell'esperienza umana e utilizzarla come materia prima da trasformare in dati comportamentali", che vengono poi commercializzati. Inizialmente, Google analizza i dati relativi ai movimenti degli utenti sulla propria piattaforma per offrire ricerche mirate in base agli interessi. Successivamente, gli amministratori comprendono che tracciare queste "briciole digitali" permette di capire gli interessi degli utenti e, in una certa misura, di prevedere il loro comportamento futuro. Il passaggio fondamentale che segna il cambiamento è il momento in cui Google decide di monetizzare queste informazioni, vendendole per creare pubblicità personalizzate per ogni utente. I DB con la mole di informazioni di dati che già possedevano, e con lo sviluppo tecnologico che ha permesso di

incrementare la facilità nella raccolta di questi ultimi, hanno iniziato ad assumere un ruolo sempre più rilevante in un settore in continua crescita.

Più che semplici rivenditori di informazioni, i data broker facilitano lo scambio di informazioni tra le società e, infine, creano mercati per i dati dei consumatori. Ciò incentiva ulteriormente la sorveglianza tra diversi tipi di entità. In questo modo, riproducono e ampliano i processi di “mercificazione del pubblico”, profondamente radicati nei processi storici di espansione capitalista.

Queste pratiche di sorveglianza sono caratterizzate da ciò che è stato definito come una “perdita asimmetrica della privacy”. Le persone sono esposte a forme di monitoraggio sempre più estese, mentre le istituzioni che effettuano il monitoraggio e le informazioni raccolte rimangono nascoste. L’asimmetria della privacy è particolarmente rilevante per il settore dei data broker, che opera da tempo senza consapevolezza pubblica o controllo normativo diretto. La privacy degli individui sotto sorveglianza è compromessa, mentre gli osservatori operano senza significativi controlli (Crain, 2017).

Il problema non è che la trasparenza non sia stata ancora configurata correttamente; è che il quadro di empowerment dei consumatori, ovvero riconoscere che il consenso del soggetto dei dati sia essenziale per il loro trattamento, fraintende le dinamiche chiave della sorveglianza commerciale e quindi offre soluzioni politiche imperfette. I data broker ottengono principalmente informazioni di seconda mano, poiché i loro clienti sono coinvolti nella raccolta delle informazioni, agendo sia come acquirenti che venditori in vari mercati. Essi rappresentano un punto cruciale nel sistema di sorveglianza, poiché si occupano di arbitrare le informazioni, acquisendole, riconfezionandole e rivendendole in contesti diversi.

I dati degli utenti e, di conseguenza gli utenti stessi, fanno parte del processo di “commodification”. La loro attività sul web costituisce un lavoro inconsapevole a vantaggio delle aziende più o meno grandi. Di conseguenza, i data broker hanno costruito un ambiente in cui gli individui sono “costantemente sorvegliati e valutati, indagati ed esaminati, categorizzati e raggruppati, valutati e classificati, numerati e quantificati, inclusi o esclusi, e, di conseguenza, trattati diversamente” (Christl, 2017). Pertanto, sono attori fondamentali del capitalismo di sorveglianza poiché si impegnano in una sorta di “arbitraggio dell'informazione” acquistando, reinterpretando, riconfezionando e vendendo dati dei consumatori in contesti diversi. Ci sono anche serie preoccupazioni che la pandemia e le tecnologie impiegate per combatterla, come i sistemi di identità digitale, i passaporti vaccinali e le app per l'attraversamento delle frontiere, abbiano effettivamente ampliato le opportunità per i data broker di raccogliere dati.

In alcuni casi è più semplice acquistare dati dettagliati su una popolazione dai data broker piuttosto che richiedere lo stesso tipo di dati al governo. Inoltre, i dati acquistati non possono essere ulteriormente esaminati o confermati a causa delle protezioni della proprietà

intellettuale dei data broker. I data broker accedono spesso gratuitamente ai dati di individui, che generalmente non sono consapevoli che i loro dati verranno riutilizzati e venduti a terzi per convincere gli utenti stessi ad acquistare prodotti che, altrimenti, non avrebbero acquistato o di cui non sarebbero venuti a conoscenza. Questo processo è molto più efficace rispetto alla pubblicità tradizionale. Rappresenta probabilmente un trasferimento piuttosto che la creazione di ricchezza.

Allo stesso modo, è estremamente difficile sfuggire alla sorveglianza dei data broker. L'acquisizione dei dati è sempre più pervasiva, soprattutto sui social media; ad esempio, dallo studio "Privacy Risks with Facebook's PII-Based Targeting: Auditing a Data Broker's Advertising Interface" (Venkatadri, 2019) è emerso che una percentuale sorprendentemente elevata di account Facebook (oltre il 90% negli Stati Uniti) è collegata con successo alle informazioni dei data broker, anche se ciò non significa che l'accuratezza delle informazioni dei DB abbia un livello di accuratezza pari. (Reviglio, 2022).

2.6 IL PROCESSO DI "COMMODIFICATION"

Per commodification si intende, in contesti sociali, il processo attraverso il quale un bene non commerciale diventa commerciale, tipicamente con connotazioni di "corruzione attraverso il commercio"; invece, il termine, commoditization è utilizzato in contesti economici per indicare quando il mercato di un prodotto esistente diventa un mercato di commodity, dove i prodotti sono intercambiabili e vi è una forte competizione sui prezzi.

È famosa la frase: "Microprocessors are commoditized. Love is commodified" (i microprocessori sono "commoditizzati". L'amore è "commodificato"). Giuridicamente ci si riferisce alla dislocazione della persona umana dalla sfera dell'indisponibilità giuridica a quella della libera disponibilità individuale (Sturabotti, 2021). In questo modo, il corpo umano e le sue parti, la sessualità, la riproduzione, così come la cultura di un popolo, l'identità personale e i dati personali, diventano beni che possono essere valutati economicamente, inseriti nel mercato e soggetti alle sue regole. La legittimità degli atti di negoziazione si basa sulla volontà liberamente espressa dal soggetto giuridico, a condizione, ovviamente, che esso sia nel pieno delle sue capacità di agire. Questo consente lo sfruttamento commerciale del bene protetto, come le informazioni personali, riconoscendolo come bene immateriale disponibile e negoziabile. Questo processo implica la trasformazione dei dati su preferenze, comportamenti e caratteristiche personali in beni di valore economico per le aziende che li utilizzano per vari scopi, tra cui, ovviamente, marketing, pubblicità e analisi di mercato. Il concetto di commodification ha radici profonde nella teoria economica e sociologica. Storicamente, l'idea può essere fatta risalire al capitalismo industriale, quando la produzione di beni è stata trasformata in un processo intensivo di sfruttamento di risorse naturali e umane. Marx è stato uno dei primi a teorizzare la commodification, descrivendo come il capitalismo trasforma lavoro e natura in merci. Nel suo lavoro, Marx ha delineato il

concetto di "feticismo delle merci", in cui le relazioni sociali tra individui sono mediate dalle relazioni tra oggetti e merci. Secondo Marx, la commodification non solo riduce la natura e il lavoro umano a beni vendibili, ma aliena anche i lavoratori dai prodotti del loro lavoro. Questo processo di alienazione implica che i lavoratori perdano il controllo sulla propria attività produttiva e sui frutti del proprio lavoro, che vengono appropriati dai capitalisti per generare profitto. Marx sosteneva che questo sistema capitalista crea un ambiente in cui i lavoratori sono visti non come esseri umani con bisogni e capacità uniche, ma come strumenti di produzione.

Nel contesto sociologico, il concetto è stato ulteriormente sviluppato da studiosi come Georg Simmel e Max Weber, che hanno esplorato come le relazioni sociali e culturali vengono trasformate in relazioni di mercato. Simmel, in particolare, ha discusso l'impatto della commodification sulle relazioni umane, sottolineando come il denaro e il mercato influenzano le interazioni sociali.

Con l'avvento della rivoluzione digitale, la commodification ha assunto una nuova dimensione. I dati personali, una volta considerati privati e intimi, sono diventati uno dei beni più preziosi nel mercato moderno. L'uso crescente dei big data e delle tecnologie di intelligenza artificiale ha ulteriormente accelerato questo processo, rendendo i dati personali facilmente commerciabili e più preziosi a livello macro ma con un valore irrisorio a livello di singoli consumatori.

- Il caso di John Moore

Uno dei casi più famosi delle conseguenze della commodification è il caso della Corte Suprema della California riguardante John Moore contro l'università della California, verificatosi nel 1990 (1990); un caso emblematico per comprendere le implicazioni giuridiche e etiche della commodification delle parti del corpo umano e dei relativi diritti di proprietà. John Moore, un paziente affetto da leucemia, subì l'asportazione della milza presso il Medical Center dell'Università della California, Los Angeles (UCLA). I medici scoprirono che i suoi tessuti producevano una particolare proteina utile per lo sviluppo di farmaci innovativi, i Linfociti T. La linea cellulare fu dapprima brevettata nel 1981, in seguito l'università e i due medici stipularono un contratto con due case farmaceutiche che brevettarono nel 1984 una serie di medicinali innovativa. Moore fece causa per conversione (per il presunto furto delle sue cellule), ma la Corte Suprema stabilì che non aveva diritto di proprietà sui suoi tessuti una volta rimossi dal corpo. Moore riuscì unicamente ad ottenere il risarcimento del danno causato dalla mancata informazione della scoperta scientifica, configurabile come violazione del rapporto fiduciario tra medico-paziente

Questo caso suggerisce la necessità di riflettere su diversi aspetti di carattere generale sul processo di commodification:

1. **Equità e Giustizia Distributiva:** come bilanciare i benefici economici derivanti dalla ricerca con i diritti degli interessati.

2. **Trasparenza e Consenso Informato:** come garantire che i pazienti siano pienamente informati e consensuali riguardo all'uso dei loro campioni biologici.

Il caso Moore evidenzia l'importanza di un quadro regolatorio che protegga i diritti dei pazienti senza ostacolare il progresso scientifico, promuovendo una giusta distribuzione dei benefici derivanti dalle innovazioni biomediche.

Quando si discute di commodification, vengono frequentemente discussi i seguenti modelli:

- il primo, noto come **zero-price**, rappresenta un sistema in cui la fornitura di dati personali è percepita come una componente implicita del servizio offerto gratuitamente dal fornitore, tipicamente una piattaforma di servizi digitali, anziché come una contropartita per il servizio utilizzato.
- il secondo, denominato **personal data economy**, vede l'utente di fronte a offerte commerciali per la fornitura di dati personali, che gli permettono di ottenere prestazioni di qualità superiore; oppure si potrebbe pensare ad una vera e propria ricompensa monetaria (Personal Data Broker).
- il terzo, chiamato **pay for privacy**, prevede che l'utente possa liberamente determinare la forma della contropartita, scegliendo tra il conferimento dei propri dati personali o il pagamento in denaro.

In ciascuno di questi modelli, il consenso ha un ruolo fondamentale, ma le condizioni al contorno per esercitarlo sono ciò che fanno la differenza per i soggetti dei dati.

2.7 POLICY PROPOSTE DALLA LETTERATURA ECONOMICA

Per quanto riguarda i modelli presi in considerazione nella sezione 1, al termine della spiegazione dell'elaborato, ogni autore ha espresso la propria proposta di regolamentazione per migliorare la competizione nel settore dei DB.

Per quanto riguarda "Selling strategic information in digital competitive markets" (David Bounie, et al., 2021), gli autori sostengono che l'accesso ai dati sia un fattore critico che influenza l'intensità della concorrenza in un qualsiasi settore a valle. Le autorità di concorrenza potrebbero garantire parità di condizioni assicurando un accesso equo ai dati alle aziende concorrenti. In secondo luogo, con lo sviluppo del big data, del machine learning e dell'intelligenza artificiale, i data broker possono segmentare con maggiore precisione la domanda dei consumatori. Così, mentre i data broker passano dalla discriminazione di prezzo di terzo grado a quella di primo grado, il benessere dei consumatori diminuisce. Pertanto, le strategie dei dati che utilizzano nuove tecnologie di micro-targeting dovrebbero essere esaminate attentamente dalle autorità di concorrenza.

Successivamente, nelle proposte di possibili estensioni del modello, si afferma che, quando due data broker competono sul mercato delle informazioni, potrebbero non trovare ottimale limitare le vendite di dati a una sola azienda, in opposizione rispetto a quanto emerge dal

lavoro stesso; tuttavia, nel mercato sembra accadere il contrario; il settore dei data broker sta diventando meno competitivo. A motivo di ciò si possono citare le acquisizioni di Bluekai e Datalogix da parte di Oracle, oppure Equifax e FICO hanno concordato di condividere informazioni sui dati finanziari dei consumatori. Ciò potrebbe cambiare gli incentivi dei data broker a vendere informazioni a tutti i partecipanti del mercato.

Infine, il meccanismo attraverso il quale le informazioni vengono vendute è un elemento importante per l'equilibrio. Questo ultimo aspetto verrà trattato nell'ultima sezione.

Per l'elaborato "Data Brokers Co-Opetition" (Gu, et al., 2021) gli autori, come analizzato in precedenza, ritengono che la condivisione dei dati possa verificarsi in quanto i DB possono attenuare la concorrenza quando i dati presentano una qualche forma di sostituibilità. In secondo luogo, la cooperazione consente ai data broker di internalizzare le inefficienze a valle, poiché gli acquirenti potrebbero essere meno efficienti dei data broker nell'unire più set di dati. A sua volta, viene identificato un possibile compromesso tra gli effetti positivi dell'internalizzazione dei costi, in linea con lo spirito della Strategia sui dati dell'UE (Commissione Europea, 2020), e gli effetti negativi della condivisione dei dati legati alla riduzione della concorrenza in questo mercato opaco.

Inoltre, l'Unione Europea e gli Stati Uniti hanno seguito approcci normativi diversi su come i dati dovrebbero essere gestiti da intermediari, terze parti e rivenditori. L'Unione Europea ha affrontato la questione della privacy in modo più rigoroso. Più nello specifico, il GDPR dell'UE ha rafforzato le condizioni per il consenso da parte dei consumatori, che devono essere esplicitamente informati sull'utilizzo finale dei dati raccolti. La condivisione dei dati tra diversi data broker senza autorizzazione dei consumatori è considerata illegale, al punto che tale regolamentazione viene spesso evocata con enfasi come la "morte dei dati di terzi". Alla luce dell'analisi, il GDPR dell'UE potrebbe avere alcuni effetti pro-competitivi "non intenzionali" nel mercato a monte dei dati. In particolare, la necessità del consenso esplicito dei consumatori per la condivisione dei dati dovrebbe ridurre la prevalenza di questa pratica, con la conseguente possibilità per le imprese a valle di trattenere parzialmente una parte del surplus generato dai dati.

Inoltre, la maggior parte dell'attenzione dei responsabili politici è stata rivolta all'uso finale dei dati e a come la condivisione dei dati possa creare esternalità positive ed effetti pro-competitivi. Tuttavia, è stata prestata poca attenzione ai dati come input, prodotti, gestiti e commercializzati dai broker di dati.

Gli autori, infine, affermano che l'analisi evidenzia che le pratiche di cooperazione dei DB potrebbero richiedere un'ulteriore attenzione da parte di un regolatore.

Nel lavoro “Data Brokers Competition, Synergic Datasets, and Endogenous Information Value” (Abrardi, et al., 2024), tenendo conto delle conclusioni, ovvero che all’equilibrio, il surplus dei consumatori:

- diminuisce con l'aumentare dell'accuratezza aggiuntiva ottenuta combinando i set di dati di DB1 e DB2,
- aumenta con l'aumentare del grado di differenziazione verticale tra DB1 e DB2, che aumenta al decrescere della differenziazione

e che è vero il contrario per il benessere totale; inoltre,

- per qualsiasi valore dell’accuratezza dei dati di DB1 e dei parametri appena descritti, invece, il benessere totale è maggiore rispetto al caso in cui i dati sono assenti,

i consumatori possono essere danneggiati dalla vendita dei dati anche se sono sub-modulari e anche se i data broker non coordinano le loro vendite. Infatti, anche sinergie relativamente deboli sono sufficienti a ridurre l'ingresso a valle e, a sua volta, il surplus dei consumatori. Infine, quando i data broker coordinano le loro vendite di dati, l'esito monopolistico del mercato risultante causa ulteriori danni ai consumatori.

Da una prospettiva di policy, i risultati implicano che garantire un “campo di gioco equo” nel mercato dei DB; cioè, ridurre il livello di differenziazione verticale nel mercato dei DB, può avere un esito positivo per i consumatori. Tale campo di gioco deve essere non solo in termini di simile accuratezza dei dati. Infatti, il potere di mercato dei DB deriva dall'aver dati esclusivi sui consumatori, come affermato anche da altri lavori, il che a sua volta aumenta la forza delle sinergie se i dataset vengono combinati. Questo può essere particolarmente dannoso quando le sinergie dei dati permettono loro di coordinare le vendite, poiché i DB possono estrarre tutto il surplus disponibile dalle aziende, aumentando a loro volta la concentrazione del mercato a valle. Gli autori concludono affermando che i risultati mettono in luce il potenziale dei DB di indirizzare la concorrenza e l'accesso ai dati e ciò richiede ulteriori indagini sulle loro pratiche di raccolta dei dati e sul funzionamento dei loro algoritmi.

L’analisi sulla figura dei personal data broker “Do Consumers Benefit from Selling Their Data? The Economic Impact of Personal Data Brokers on Digital Markets” (Krämer, et al., 2018) si pone come domanda di ricerca se i consumatori stiano meglio con un PDB nel mercato. Ciò sembra essere il caso solo quando il PDB è molto efficiente, il che probabilmente potrebbe non accadere nella pratica. Questo pone la questione di come potrebbe essere creata una regolamentazione ad hoc, se possibile, per migliorare le condizioni sotto le quali un PDB avvantaggia i consumatori. Di seguito sono riportati i tre interventi specifici ipotizzati e il loro impatto sui risultati del mercato e sul surplus dei consumatori.

Le conclusioni possono essere prese in considerazione per il settore generale dei data broker.

1. Regolamentazione sulla portabilità dei dati

Gli autori affermano che, nella pratica, i decisori politici hanno una certa discrezionalità nel definire l'ambito dei dati soggetti alla portabilità dei dati. Ad esempio, si possono distinguere

- dati volontari: dati esplicitamente forniti dagli utenti, come i "Mi piace",
- dati osservati: dati forniti implicitamente dagli utenti, come clickstream o geolocalizzazione e
- dati derivati: dati generati mediante ricombinazione o analisi statistica.

Nelle sue linee guida interpretative del GDPR, il Comitato europeo per la protezione dei dati (EDPB) suggerisce che l'ambito dei dati portabili ai sensi dell'Articolo 20 dovrebbe includere i dati volontari, ma escludere i dati derivati (Commissione Europea, 2017). Tuttavia, rimane qualche incertezza riguardo alla misura in cui i dati osservati sono soggetti ai diritti di portabilità dei dati. Di conseguenza, l'EDPB potrebbe rivedere le sue linee guida e aumentare o ridurre l'ambito dei dati portabili. Inoltre, i responsabili politici potrebbero aumentare la frequenza con cui i dati possono essere trasferiti, permettendo la portabilità dei dati in tempo reale anziché solo una tantum come richiesto dal GDPR. Infatti, ai sensi dell'Articolo 6, paragrafo 9 del Digital Markets Act europeo, gli utenti delle piattaforme online molto grandi hanno il diritto alla portabilità continua e in tempo reale dei dati. Infine, potrebbero essere promulgate nuove normative per consentire ai consumatori di trasferire non solo i dati personali, ma anche i dati non personali. Tale regolamentazione è prevista nell'UE nel contesto dei prodotti connessi (IoT) ai sensi del Data Act.

Gli autori ottengono soluzioni contrastanti: da un lato, si mostra che con una quota maggiore di portabilità dei dati, l'equilibrio di reddito positivo emerge già per livelli inferiori di efficienza del PDB. Inoltre, la soglia di efficienza del PDB, al di sopra della quale i consumatori traggono vantaggio dal PDB, diminuisce con la quota di portabilità dei dati. D'altro canto, come si deduce anche intuitivamente, una maggiore portabilità dei dati diminuisce l'investimento in qualità del CSP nell'equilibrio di reddito minimo. Appare utile ricordare a questo punto le definizioni già introdotte nella sezione 1: per equilibrio di reddito minimo (minimum income equilibrium) si intende la ricompensa minima offerta dal PDB ai consumatori, tale che sia sufficiente a garantirne la partecipazione, ma che non genera effetti sufficienti per il surplus del consumatore. Per equilibrio di reddito positivo (positive income equilibrium) si intende, invece, una ricompensa che risulti positiva e significativa ai consumatori.

Di conseguenza, l'ampliamento della portata della portabilità dei dati risulta vantaggioso per i consumatori solo se il PDB è sufficientemente efficiente nel generare entrate dai dati, permettendo così il raggiungimento di un equilibrio di reddito positivo. Invece,

nell'equilibrio di reddito minimo, una maggiore portabilità dei dati peggiora sempre la condizione dei consumatori.

2. Regolamentazione "do-not-sell-my-data"

Successivamente, viene considerato un regolamento Do-Not-Sell-My-Data (DNS) che garantisce ai consumatori il diritto di richiedere al CSP di non vendere i propri dati personali a terzi. Questo intervento politico si ispira al California Consumer Privacy Act, che impone alle aziende basate sui dati di offrire ai consumatori la possibilità di rinunciare alla vendita dei propri dati personali. Precisamente il CCPA (Cal. Civ. Code § 1798.135) afferma che un'azienda che rientra nel CCPA deve "fornire un collegamento chiaro e ben visibile sulla home page Internet dell'azienda, intitolato "non vendere le mie informazioni personali", a una pagina Web Internet che consente a un consumatore, o a una persona autorizzata dal consumatore, di rinunciare alla vendita delle informazioni personali del soggetto interessato."

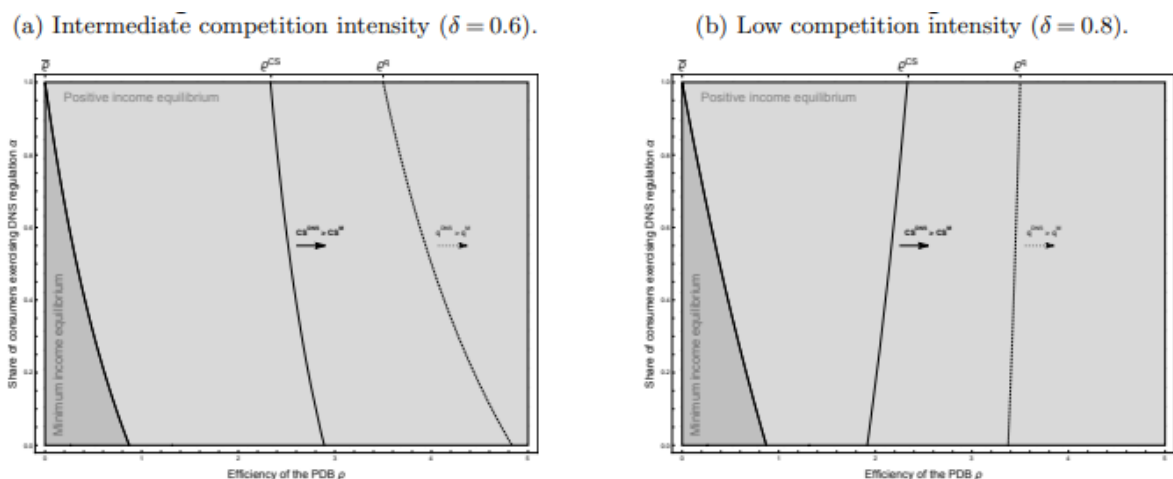


FIGURA 19: Risultati di mercato con e senza un PDB sotto la regolamentazione DNS a diversi livelli di competizione (Krämer, et al., 2018)

La Figura 18 illustra i risultati di mercato con e senza un PDB sotto regolamentazione DNS, in dipendenza dell'efficienza del PDB e della quota di consumatori che esercitano la regolamentazione DNS. Come nel caso base, i consumatori possono trarre beneficio dal PDB sotto regolamentazione DNS solo nell'equilibrio di reddito positivo e sono sempre in una condizione peggiore rispetto al caso senza PDB nell'equilibrio di reddito minimo. Se l'intensità della competizione è relativamente alta, una maggiore quota di consumatori che optano per DNS rende più probabile che i consumatori traggano beneficio dal PDB, poiché ciò abbassa la soglia critica di efficienza. Questo effetto sulla soglia % può essere osservato nel pannello di sinistra della Figura 18. Al contrario, se l'intensità della competizione è bassa,

l'effetto è inverso e la soglia critica aumenta con una maggiore quota di consumatori che optano per DNS. Questo effetto sulla soglia critica % CS può essere osservato nel pannello di destra della Figura. Ciò avviene perché, quando l'intensità della competizione è debole, la perdita relativa di ricavi dai dati per il CSP a causa della regolamentazione DNS è maggiore e, di conseguenza, il CSP riduce ulteriormente i suoi investimenti in qualità rispetto a quando la competizione è forte il che alla fine danneggia i consumatori.

3. Regolamentazione con campagne di (in)formazione

L'ultimo possibile intervento politico esaminato dagli autori sono le campagne di informazione che formano i consumatori sul loro diritto di trasferire i dati dai CSP e migliorano la loro capacità di avviare un trasferimento in un PDB. Ciò dovrebbe aumentare l'adozione di un PDB, che a sua volta fornisce più dati e maggiori entrate derivanti dai dati. In effetti, queste sono relativamente simili al caso esaminato in precedenza "regolamentazione sulla portabilità dei dati", poiché una maggiore adozione del PDB comporta anche un aumento della quantità di dati utente trasferiti dal CSP al PDB, quindi, una maggiore conoscenza coinciderebbe con una portabilità maggiore (come si può intuire anche intuitivamente). Un aumento della quota di consumatori informati ed esperti di tecnologia aumenta gli effetti negativi del PDB sulla qualità del CSP e sul surplus dei consumatori nell'equilibrio di reddito minimo.

In conclusione, le campagne di informazione che determinano una maggiore adozione del PDB sono vantaggiose solo per quei consumatori che lo adottano e se il PDB è solo se è sufficientemente efficiente nel generare entrate derivanti dai dati, in modo tale da far emergere un equilibrio di reddito positivo.

2.8 I CASI KOCHAVA E X-MODE

I seguenti due casi dell'antitrust americano riguardano la vendita di dati di geolocalizzazione con timestamp da parte di due data broker, rispettivamente Kochava (2022) e X-Mode (2018-2020); le considerazioni sono simili, in quanto, il modello di business dei due DB, per il quale la FTC ha portato avanti il caso, era lo stesso.

Kochava

La Federal Trade Commission (FTC) ha intentato una causa contro il data broker Kochava Inc. Società con sede nell'Idaho, che acquista vaste quantità di informazioni sulla posizione di dispositivi mobili per la vendita di dati di geolocalizzazione che possono essere utilizzati per tracciare gli spostamenti delle persone da e verso luoghi sensibili. I dati che Kochava ha venduto ai suoi clienti, oltre a contenere informazioni sensibili, possono mostrare le visite delle persone, ad esempio, a cliniche per potenziali genitori (ginecologiche o di aborto),

luoghi di culto, rifugi per senza tetto, centri per vittime di violenza domestica, e strutture di recupero da dipendenze.

Kochava ha affermato di offrire "ricchi dati geografici che abbracciano miliardi di dispositivi in tutto il mondo". Ha inoltre sostenuto che il suo flusso di dati sulla posizione "fornisce dati grezzi di latitudine/longitudine con volumi di circa 94 miliardi di transazioni geografiche al mese, 125 milioni di utenti attivi al mese e 35 milioni di utenti attivi al giorno, in media osservando più di 90 transazioni giornaliere per dispositivo". I dati di geolocalizzazione che includono le coordinate con timestamp che mostrano la posizione dei dispositivi mobili nel tempo; In particolare, ai loro dispositivi mobili è associato un "identificatore individuale e costante" che permette di tracciare i movimenti dei consumatori nell'arco di un giorno, settimana, mese e persino anno, anche in luoghi sensibili e personali. Kochava promette ai suoi clienti che i dati sono così precisi da collocare accuratamente i movimenti dei consumatori entro pochi metri, abbastanza per dire non solo in quale edificio si trovino i consumatori, ma anche in quale stanza. L'identificatore è noto come Mobile Advertising ID (MAID). Un MAID è assegnato dal sistema operativo di un dispositivo mobile per consentire alle aziende di tracciare l'attività mobile di un consumatore e viene utilizzato per inviare pubblicità mirate. La FTC afferma che Kochava stessa assicura che i MAID non offrano alcuna anonimizzazione per i consumatori, che vende dati che collegano direttamente i MAID alle informazioni identificative individuali dei consumatori, e che incoraggia espressamente i suoi clienti ad utilizzare questi dati. Infatti, Kochava ostenta la capacità di collegare i veri nomi, gli indirizzi, gli indirizzi e-mail e i numeri di telefono dei consumatori a informazioni sensibili, tra cui il loro genere, stato civile e età, affermando che: "a partire dai più di 320 milioni di visitatori unici mensili, creare profili di persone e ottenere nomi, indirizzi, stato civile, livello di istruzione, stabilità economica, reddito annuo, indirizzi email, numeri di telefono, genere ed età". Di conseguenza, i clienti di Kochava non hanno neppure bisogno di attuare inferenze o compiere analisi aggiuntive.

I dati precisi di geolocalizzazione di Kochava possono essere utilizzati per identificare i consumatori che hanno visitato una clinica per aborti e, di conseguenza, possono aver avuto o considerato di avere un aborto. Infatti, solo nei dati resi disponibili da Kochava, il ricorrente ha identificato un dispositivo mobile che ha visitato una clinica per la salute riproduttiva delle donne e ha tracciato tale dispositivo mobile fino a una residenza monofamiliare. I dati possono anche essere utilizzati per identificare professionisti medici che eseguono o assistono nell'esecuzione dei servizi di aborto. Come altro esempio, i dati possono essere utilizzati per tracciare i consumatori fino ai luoghi di culto, rivelando così le credenze e le pratiche religiose dei consumatori. Infatti, il ricorrente ha identificato nel campione di dati di Kochava dispositivi mobili che erano situati in luoghi di culto ebraici, cristiani, islamici e di altre denominazioni religiose. Come ulteriore esempio, i dati possono essere utilizzati per tracciare i consumatori che hanno visitato rifugi per senzateetto, centri

per vittime di violenza domestica o altre strutture rivolte a elementi della popolazione a rischio, e come ultimo esempio, la posizione di un dispositivo mobile di notte si potrebbe associare all'indirizzo di casa dell'utente e potrebbe essere combinata con i dati relativi alla proprietà per scoprirne l'identità. In effetti, il DB ha pubblicizzato l'identificazione delle famiglie come uno dei possibili usi dei suoi dati. Queste informazioni possono rivelare la posizione dei consumatori che stanno sfuggendo alla violenza domestica o ad altri crimini. Inoltre, poiché i dati di Kochava consentono ai suoi clienti di tracciare i consumatori nel tempo, i dati potrebbero essere utilizzati per identificare anche le condizioni passate dei consumatori.

Nel database di Kochava, ai dispositivi sono collegate le seguenti informazioni:

- Nome
- Indirizzo
- E-mail
- Numero di telefono
- MAID
- Etnia
- Identità di genere
- Status di minore
- Data di nascita
- Stato civile
- "associazione politica"
- Status di genitore e numero di figli

Kochava ha venduto nei pacchetti offerti ai propri clienti le informazioni sopra elencati, ciò viola la privacy dei consumatori.

Kochava raccoglie, utilizza e vende informazioni dettagliate su ciò che i consumatori fanno sui loro dispositivi mobili attraverso App Graph (come l'ID graph di Oracle) e l'Attività dell'App (collettivamente, l'"App Graph" di Kochava). Si basa su una combinazione di dati acquisiti da altri DB e informazioni che Kochava raccoglie sull'uso delle app da parte dei consumatori attraverso il proprio Free App Analytics Software Development Kit (FAA SDK). Il FAA SDK è un insieme di strumenti digitali per lo sviluppo software raccolti in un'unica confezione, o "kit". Gli sviluppatori di app possono installare il FAA SDK nelle proprie app per facilitare il tracciamento delle inserzioni pubblicitarie e delle attività dei consumatori nell'app. Nel momento in cui un consumatore utilizza un'app mobile in cui è stato inserito l'SDK FAA, esso trasmette informazioni a Kochava sull'attività dell'utente sull'app.

Le informazioni vendute da Kochava tramite App Graph includono: il nome dell'app in uso, la data di utilizzo, quanto tempo, in secondi, viene utilizzata, il tipo di azioni che gli utenti compiono e quanto denaro l'utente spende nell'app. Come per i siti web, anche le app da

cui Kochava può estrarre informazioni vengono utilizzate per scopi privati o sensibili. Per esempio, ci sono app di incontri, app utilizzate da determinati gruppi religiosi, come app di preghiera musulmana e app che forniscono informazioni su sindromi, come cancro o malattie sessualmente trasmissibili.

Kochava segmenta i consumatori sui loro "interessi e comportamenti". e spiega che questi segmenti sono costruiti sulla base di una " Content Taxonomy (Tassonomia dei contenuti)" online. Questa tassonomia include informazioni su: "Salute riproduttiva", "cancro", "Salute delle donne", "divorzio", "lutto", "assistenza agli anziani", "adozione e affidamento", "bambini con bisogni speciali", "condizioni sessuali", "gravidanza", "vaccini", "Giudaismo" e "Islam".

La FTC afferma, che le informazioni di cui Kochava dispone forniscono un quadro completo della vita privata dei consumatori, sia online che offline, che non può essere ottenuto tramite osservazioni fisiche negli spazi pubblici. Fornisce una visione senza precedenti delle azioni personali, delle decisioni e dei comportamenti di un consumatore; accusando la società di invadere le aree più private della vita dei consumatori e causare danni sostanziali ai consumatori; l'antitrust americano infine accusa Kochava di causare, oltre a violare la privacy dei consumatori, altre forme di danno ai consumatori, come stigma, discriminazione, violenza fisica, angoscia emotiva.

Attraverso i dati precisi di geolocalizzazione di Kochava, chi compra i dati è in grado di conoscere i consumatori che, ad esempio, hanno visitato luoghi sensibili, esponendoli a questi danni aggiuntivi. Inoltre, la probabilità che tali danni aggiuntivi avvengano è aggravata dalla mancanza di controlli da parte di Kochava su chi accede a questi dati e su come tali entità li utilizzano. Kochava, infatti, viene anche accusata di vendere i propri dati a clienti che hanno fornito un numero irrisorio di informazioni sulla propria azienda.

Per evidenziare i rischi delle azioni commesse da Kochava, la FTC porta, a favore della sua tesi, un caso paragonabile: un'altra società ha pubblicizzato la capacità di raggiungere "donne vulnerabili all'aborto" e di mostrare loro, utilizzando i MAID dei dispositivi delle donne in questione, incluso sui loro feed Facebook, Instagram e altri social media, pubblicità mirate. Secondo i report giornalistici, uno di questi annunci recitava: "Hai preso la prima pillola presso la clinica? Potrebbe non essere troppo tardi per salvare la tua gravidanza." Secondo i report, gli annunci indirizzavano i consumatori a siti web che cercavano di persuadere i consumatori a tentare una procedura di "reversibilità dell'aborto" non supportata scientificamente.

X-Mode

X-Mode Social, ora conosciuta come Outlogic, è un DB specializzato nella raccolta di dati di localizzazione da applicazioni mobili. Fondata nel 2013 e con sede a Reston, Virginia, Outlogic fornisce dati di localizzazione precisi per vari scopi commerciali. La società si

distingue per l'analisi e la fornitura di dati che supportano decisioni strategiche in settori come l'immobiliare e il finanziario. Le peculiarità di questo caso sono le seguenti:

- attraverso le proprie app, app-partner e altri data broker, X-Mode ha raccolto quotidianamente oltre 10 miliardi di punti dati di localizzazione da tutto il mondo. X-Mode sostiene che questi dati di localizzazione sono accurati al 70% entro un raggio di 20 metri o meno;
- inoltre, la FTC afferma che dal 2013 circa, il sistema operativo Android ha incluso un controllo privacy che permette agli utenti di "disattivare la personalizzazione degli annunci." Questo controllo della privacy permette ai consumatori di impedire ai marketer di utilizzare i MAID dei loro telefoni per creare profili sui consumatori o mostrare loro annunci personalizzati. Quando i consumatori attivavano questo controllo sui loro telefoni, il sistema operativo Android trasmetteva un MAID del telefono a un'app, insieme ad altre informazioni richieste, e includeva una "flag" che informava l'app della scelta dei consumatori di disattivare la pubblicità personalizzata. Dal giugno 2018 fino a luglio 2020, X-Mode ha raccolto i MAID, i dati di localizzazione mobile e le "flag" dei consumatori che avevano disattivato la pubblicità mirata sui loro telefoni Android e, in molti casi, contrariamente alle scelte di privacy di questi consumatori, ha fornito l'accesso a questi dati ai marketer e ad altri clienti. X-Mode non ha implementato le necessarie salvaguardie tecniche e la supervisione per garantire che le scelte di privacy dei consumatori attivate sui loro telefoni Android fossero rispettate e che i loro dati di localizzazione non fossero più raccolti o venduti per scopi di pubblicità personalizzata.

La FTC ha accusato i due DB dei seguenti violazioni:

1. Vendita sleale di dati sensibili
2. Raccolta e utilizzo sleali dei dati sulla posizione dei consumatori
3. Raccolta e utilizzo sleali dei dati sulla posizione dei consumatori senza verifica del consenso
4. Categorizzazione sleale dei consumatori basata su caratteristiche sensibili per scopi di marketing

2.9 DB E GEOPOLITICA, CASO EQUIFAX E TIKTOK

Il caso della violazione dei dati di Equifax ha rilevato punti di connessione tra DB e stabilità geopolitica. Sul sito della FBI, in riferimento al suddetto caso, si legge: "Oggi (10 febbraio 2020), il dipartimento di giustizia degli Stati Uniti ha annunciato accuse contro quattro hacker supportati dall'esercito cinese in connessione con l'attacco informatico del 2017 contro Equifax [...]. L'intrusione ha portato al più grande furto noto di informazioni personali identificabili mai realizzato da attori sponsorizzati dallo stato. [...] il malware ha

permesso agli hacker di raccogliere indirizzi, date di nascita, numeri di previdenza sociale e altri dati su circa 145 milioni di americani. L'atto d'accusa di oggi accusa i membri dell'Esercito Popolare di Liberazione, le forze armate della Repubblica Popolare Cinese, di essere dietro a quell'attacco malware" (FBI, 2020).

Alla luce dell'avanzamento delle tecnologie di sorveglianza, diversi governi, tra cui l'Unione Europea e la Cina, hanno introdotto misure che promuovono la localizzazione dei dati o restrizioni sul flusso transfrontaliero dei dati, citando motivazioni di sicurezza nazionale o privacy. Tuttavia, la sola localizzazione dei dati non può prevenire le violazioni della privacy e talvolta aumenta l'accesso dei governi nazionali a informazioni personali sensibili. Inoltre, restrizioni eccessivamente ampie sul flusso dei dati possono interrompere le comunicazioni e il commercio online, specialmente considerando che quasi tutte le imprese moderne gestiscono siti web o app, conducono transazioni globali e dipendono in vari modi dal flusso internazionale dei dati.

Mentre gli Stati Uniti valutano diverse opzioni per affrontare i rischi per la sicurezza nazionale derivanti dalle app mobili come TikTok e dai data broker esteri, è importante considerare come restrizioni generalizzate sul flusso transfrontaliero dei dati potrebbero creare conseguenze indesiderate a lungo termine. La localizzazione dei dati potrebbe spingere altri paesi a istituire limitazioni reciproche nei confronti degli Stati Uniti o a indebolire le negoziazioni globali sui flussi di dati liberi in futuro. Le imprese statunitensi hanno già sperimentato una significativa incertezza dopo le decisioni della Corte di Giustizia dell'Unione Europea nei casi Schrems I e Schrems II, che minacciavano di ridurre i flussi di dati transatlantici. Quando la Francia si è unita agli Stati Uniti nel bloccare TikTok sui dispositivi emessi dal governo nel marzo 2023, ha anche aggiunto app statunitensi come Twitter e Netflix a tale elenco. Se si può presentare il rischio che anche alleati politici ed economici stretti come l'Unione Europea e gli Stati Uniti non possano mantenere la fiducia nei flussi di dati, allora ulteriori normalizzazioni della localizzazione dei dati potrebbero causare un collasso del commercio e delle comunicazioni nell'economia globale. Inoltre, anche se la localizzazione forzata dei dati per TikTok o per future app mobili non statunitensi potrebbe rallentare i governi stranieri dall'accedere a quei database specifici, potrebbe contemporaneamente agevolare l'accesso o un possibile abuso da parte del governo statunitense.

Tutti i data broker e le aziende private statunitensi dovrebbero seguire determinati principi di pratiche di informazione corretta: riduzione della raccolta dei dati a scopi specifici, cancellazione dei dati non appena non sono più necessari, permettendo una maggiore trasparenza e sorveglianza, e altro ancora. (Center for Strategic and International Studies (CSIS), 2023)

Benché TikTok non rientri nella definizione di DB della sezione 1, è interessante notare che estragga informazioni sensibili e non sui propri consumatori (come un first-party DB) e possa essere considerato come un fornitore di dati per il governo cinese. Inoltre, di recente

il governo statunitense ha deciso di proseguire l'iter per l'esclusione della holding cinese di TikTok USA. La Commissione per l'Energia e il Commercio della Camera degli Stati Uniti ha approvato all'unanimità il disegno di legge all'inizio di marzo del 2024. Tale disegno di legge richiede a ByteDance (Pechino), la società che controlla TikTok, di vendere TikTok USA agli investitori occidentali entro un periodo di 180 giorni, altrimenti avrebbe rischiato il divieto dell'app nel Paese. Le preoccupazioni per la sicurezza nazionale sono aumentate, poiché i funzionari statunitensi temono che Pechino possa accedere ai dati personali dei cittadini americani di TikTok USA, che conta oltre 170 milioni di utenti, a causa delle leggi cinesi che impongono alle società nazionali di fornire informazioni al governo. Nonostante TikTok abbia intrapreso azioni come il "Progetto Texas" per isolare i dati degli utenti statunitensi dalla Cina, con un investimento di 1.5 miliardi di dollari, i provvedimenti dell'azienda non sono riusciti a placare le preoccupazioni del Congresso. Alla fine di aprile del 2024, il Senato degli Stati Uniti ha approvato una proposta di legge che potrebbe portare al blocco di TikTok sul territorio degli USA. Il testo è stato firmato dal presidente Biden, diventando legge. Tuttavia, l'attuazione della legge richiederà più dei 180 giorni previsti inizialmente. Il testo richiede infatti a ByteDance di cedere il social network a una società americana o di cessare le sue attività sul suolo americano, concedendo un periodo di 270 giorni.

Nove mesi rappresentano un periodo esteso, e potrebbero essere aumentati di altri tre mesi "qualora ci fossero evidenti progressi in direzione di un accordo". La prima bozza parlava di sei mesi, quindi il prolungamento a nove mesi indica che una prima decisione su questo verrà presa a gennaio. Tuttavia, questa decisione non spetterà all'attuale presidente, ma al successore. Inoltre, è probabile che ByteDance e TikTok USA avvieranno ricorsi legali, come già confermato dalla compagnia, rendendo prematuro parlare di blocco di TikTok negli Stati Uniti, almeno fino al 2025. (Biden ha firmato la legge sul blocco di TikTok negli Stati Uniti: ora che succede?, 2024)

SEZIONE 3 – DB, PREZZI, CLIENTI, CONSUMATORI E DATASET

3.1 STRUTTURE DI PREZZO NEL MERCATO DEI DATI

Tecnicamente i data broker "affittano" o "concedono in licenza" i dati; tuttavia, queste pratiche possono essere assimilate ad una vendita. Avast offre, inoltre degli esempi di clienti dei DB, alcuni già discussi in precedenza:

- altri data broker,
- inserzionisti che acquistano "informazioni di mercato" per rivolgersi a potenziali clienti,
- partiti per campagne politiche, che acquistano dati per indirizzare i propri messaggi elettorali,

- istituti finanziari che desiderano informazioni su una persona specifica prima di concederle un prestito,
- proprietari che vogliono verificare i potenziali inquilini,
- potenziali datori di lavoro che esaminano candidati per un posto di lavoro.

Dal sito di Avast si legge: “chiunque abbia \$ 20 e voglia scoprire i tuoi segreti”. (Avast, 2024)

Come definito nella sezione 2, il processo di commodification ha reso i dati del singolo utente merci non differenziabili e di poco valore; tuttavia, il valore aggregato di milioni di dati permette la differenziazione e di conseguenza la segmentazione verticale (o della qualità). Assume un significato reale, dunque, parlare di diverse qualità dei dati che possono essere venduti a prezzi diversi. I consumatori preferiscono ovviamente dati di qualità superiore rispetto a dati di qualità inferiore, a parità di prezzo. Per soddisfare le diverse esigenze dei consumatori, i produttori offrono solitamente varie linee di prodotti con diversi livelli di qualità. In generale, i DB devono considerare sia il costo marginale, sia il costo dipendente dalla qualità per prendere decisioni sulla differenziazione dei prodotti.

Per la produzione di beni fisici, il costo di riproduzione è parte del costo primario da considerare. Al contrario, per le produzioni di beni digitali, in particolare quella dei dati, il costo di riproduzione è relativamente limitato e consiste in: controllo delle versioni (aggiornamento dei dati), verifica dell'integrità, manutenzione e altro.

I dataset di record climatici grezzi o pre-elaborati possono essere riconfezionati in vari livelli di prodotti utilizzando diverse precisioni, frequenze temporali e altri parametri. Questi dataset, con prezzi e caratteristiche diversi, possono soddisfare varie esigenze dei consumatori.

Nel lavoro “A Survey on Big Data Market: Pricing, Trading and Protection” (Liang, et al., 2018) vengono analizzate le strutture di pricing nel mondo dei dati; è necessaria tuttavia, una premessa: come è noto, sono le strutture di mercato a determinare il prezzo nel commercio di beni fisici, e lo stesso vale per il mercato dei dati, in cui tali strutture influiscono ugualmente sul prezzo finale che addebiteranno i DB ai clienti.

La vendita di dati non avviene unicamente da parte dei DB, ma, in generale, vi sono diverse categorie di dati che non rientrano nelle categorie descritte nella sezione 1 che vengono vendute dai DB.

Nelle strutture di monopolio, lo stesso dataset viene analizzato, prodotto, modificato dallo stesso fornitore. In altre parole, le varie qualità e livelli di beni sono prodotti a partire dallo stesso dataset originale. Comunemente, invece di fissare un unico prezzo, i monopolisti stabiliscono diverse fasce di prezzo a seconda delle qualità e dei dettagli delle informazioni per soddisfare le esigenze dei consumatori a diversi livelli.

In condizioni di concorrenza, invece, la maggior parte dei DB dovrebbero perdere la propria posizione di mercato, e solo pochi sopravvivono. Questi “vincitori” controllano per lo più le risorse di mercato, risultando in una struttura di oligopolio. In particolare, nel mercato dei dati, i dataset originali appartengono solo a pochi proprietari. Pertanto, i proprietari dei dati hanno un forte potere di controllo sul processo di machine learning e data mining, sui prezzi di mercato, sulla concorrenza e sulle opportunità. In questo modo, i proprietari dei dati originali hanno la possibilità di catturare maggior valore.

Nel mercato dei beni, non basta considerare unicamente i costi per la determinazione del prezzo di un determinato bene; ciò vale ancora di più nel mercato dei dati. In realtà il costo è solo uno dei fattori per la determinazione del prezzo.

Secondo lo studio, possono essere categorizzate le strategie di pricing esistenti in diversi gruppi; tuttavia, non vi sono evidenze sulle strutture di prezzo applicate dai DB. In generale, possono essere presi in considerazione i modelli di pricing studiati dalla letteratura sul mercato dei dati.

I principali fattori che determinano il prezzo dei dati, esogeni al mercato dei dati stesso, sono: il costo di raccolta dei dati, il costo di analisi dei dati, il costo di gestione dei dati e la domanda dei clienti. Bisogna anche tenere in considerazione la diversità dei dati: per vendere dataset, i DB elaborano i dati grezzi al fine di soddisfare varie richieste. Questo approccio solleva una serie di questioni complesse per la valutazione dei prezzi. Ad esempio, un dataset originale deve essere riprodotto e suddiviso in diversi livelli di vari volumi, precisioni e tipi. Quindi, come valutare il prezzo rimane una questione impegnativa.

Esistono diverse strategie di prezzo per valutare il costo in base ai diversi scenari.

1. Vi sono strategie basate meramente su principi economici e,
2. Strategie basate sulla teoria dei giochi, come quelle analizzate nella sezione 1. In tali strategie, il modello considera che il prezzo sia influenzato dinamicamente dalla competizione.

Per quanto riguarda i modelli basati sui principi economici, si parla di:

- Modello di costo: è il più semplice, considera il costo totale per qualsiasi bene e stabilisce una percentuale del costo totale come profitto. Assumendo che I siano i ricavi desiderati, C il costo totale e p la percentuale di profitto. L'equazione $I = C(1 + p)$ rappresenta la relazione tra costo e reddito.
- Modello di domanda e offerta.
- Modello di prezzi differenziali: per soddisfare le varie richieste di beni, i DB potrebbero offrire diverse merci con caratteristiche (qualità) differenti. Il modello di prezzi differenziali considera le differenze tra i dati di qualità differente e propone

prezzi diversi. Ad esempio, un pacchetto di dati ad alta precisione dovrebbe avere un prezzo più alto rispetto a un pacchetto di dati a bassa precisione.

3.2 STRUTTURE DI PREZZO UTILIZZATE DAI DB

Analizzando empiricamente il mercato pubblico sul quale si possono trovare i servizi offerti dai data broker, si trova che alcuni dataset dei principali DB come Acxiom, Experian, LiveRamp, Enigma sono presenti sulla piattaforma "AWS marketplace".

AWS Marketplace è una piattaforma online per software e servizi cloud, offerta da Amazon Web Services (AWS), nella quale gli utenti possono trovare applicazioni, dati e servizi di machine learning, preconfigurati per funzionare sull'infrastruttura AWS. Ciò li agevola nel trovare, installare e gestire le soluzioni di cui hanno bisogno, senza doverle configurare da zero.

SUBSCRIPTIONS

Su AWS marketplace i DB vendono dataset sotto iscrizione/abbonamento: in questo caso offrono dati tramite abbonamenti periodici (mensili, annuali, ecc.), dove l'acquirente paga una tariffa ricorrente per accedere ai dati. I dataset vengono aggiornati periodicamente. Ciò garantisce un flusso di entrate costante ai broker e una fornitura continua di dati agli acquirenti.

FREEMIUM

In alcuni casi, i data broker possono utilizzare un modello freemium, dove una parte dei dati è offerta gratuitamente, mentre l'accesso ai dati più dettagliati o di maggiore qualità è a pagamento. Su AWS, alcuni DB offrono alcuni sample di dataset dove sono presenti solo le variabili e non record reali, oppure sono presenti solo record fittizi. Questi sample possono essere intesi come i "dictionary". Per ottenere i dataset con i dati reali bisogna "abbonarsi" al servizio.

CUSTOM PRICING

Infine, i data broker possono offrire prezzi personalizzati in base alle esigenze specifiche dell'acquirente, alla quantità di dati richiesti, e alla durata del contratto. Questo approccio è comune nelle transazioni B2B dove le esigenze possono variare notevolmente. Indiscutibilmente viene applicata discriminazione sui singoli clienti che richiedono certe tipologie e volumi di dati.

I dataset verranno approfonditi nel dettaglio nei prossimi paragrafi.

In conclusione, si può affermare che i DB vendono la maggioranza dei propri dati a diversa granularità, qualità e quantità tramite consulenze individuali i cui appuntamenti vengono presi direttamente dal sito dei DB stessi. Ciò fa presumere che riescano a identificare il

cliente (ed i clienti del cliente) con elevata precisione, offrendo il prezzo che permette loro di massimizzare il profitto.

3.3 DB E CLIENTI

I DB hanno informazioni sui propri clienti e le utilizzano per offrire il miglior prodotto al miglior prezzo. Per quanto riguarda, invece, le pratiche di screening, contrattazione e monitoraggio dei clienti da parte dei DB, esse variano a seconda del tipo di servizio (marketing, rilevamento delle frodi, credit scoring, ...), del tipo di dati forniti dal servizio (informazioni anagrafiche, sulla proprietà, sullo stile di vita, ...) o del tipo di cliente (ad esempio, istituto finanziario o rivenditore). Quando compilano informazioni governative e altre informazioni pubblicamente disponibili per fornire prodotti di ricerca delle persone al pubblico generale, i data broker tendono a effettuare un minimo screening o monitoraggio dei loro clienti. I clienti, inclusi i consumatori individuali, per acquistare i servizi offerti dai DB accedono ai loro siti web e i data broker non chiedono lo scopo per il quale il cliente utilizzerà il prodotto, come per il caso di Kochava discusso nella sezione 2.

I data broker dichiarano di proibire determinati utilizzi dei dati tramite i “termini di utilizzo” dei loro siti web, che spesso includono divieti sull'utilizzo dei prodotti per scopi illeciti, per scopi che seguono il Fair Credit Reporting Act (FCRA), che regola le modalità di gestione delle informazioni creditizie dei consumatori da parte delle aziende, o entrambi. Alcuni DB pubblicano semplicemente i loro termini di utilizzo senza richiedere ai clienti di accettarli esplicitamente; altri richiedono ai clienti di accettarli esplicitamente al momento del completamento dell'acquisto, anche se i termini potrebbero non essere visualizzati durante tutto il processo di acquisto.

Nel report “DATA BROKERS A Call for Transparency and Accountability” (Federal Commission Trade (FTC), 2014), la FTC afferma che i DB generalmente:

1. non rivedono, ovvero non esaminano nuovamente l'uso del servizio da parte del cliente;
2. non monitorano;
3. non conducono operazioni di audit: non conducono verifiche o ispezioni formali sull'uso del servizio da parte del cliente;
4. non valutano: non analizzano o giudicano come il cliente stia utilizzando il servizio.

Questo implica che, una volta venduto il servizio, i data broker non seguono da vicino come viene utilizzato dal cliente, né intervengono per assicurarsi che l'uso sia conforme a determinati standard o regolamenti, salvo eventuali disposizioni contrattuali o legali specifiche.

Alcuni data broker, invece, effettuano un certo screening e monitoraggio dei loro clienti. Il processo di screening può includere incontri o colloqui con potenziali clienti, affidarsi alla reputazione ben consolidata del potenziale cliente e condurre alcune ricerche sulla legittimità dell'attività del potenziale cliente, come la verifica dell'indirizzo aziendale, ricerche su Internet e la revisione del sito web del potenziale cliente.

Altri effettuano uno screening e un monitoraggio più significativi. Oltre al processo di screening descritto sopra, possono includere un questionario di accreditamento per determinare se il cliente è un'entità legittima e ha un uso legittimo per il prodotto, ispezioni del sito, revisioni di sicurezza, revisioni del sito web e controlli delle referenze.

Indipendentemente dal fatto che effettuino o meno lo screening o il monitoraggio dei loro clienti, i data broker che offrono servizi per la mitigazione del rischio e il marketing stipulano contratti scritti e firmati con i loro clienti che descrivono gli utilizzi consentiti e proibiti del prodotto. Gli utilizzi proibiti possono includere il riutilizzo o la rivendita dei dati senza permesso; di conseguenza, sarebbe come se fossero i DB a rivendicare il diritto di revoca dei dati.

3.4 DB E CONSUMATORI

I soggetti che volessero avere accesso ai propri dati sui database dei DB possono richiedere ai DB stessi di esercitare i propri diritti, discussi precedentemente, o semplicemente di visualizzare i dati che i DB possiedono su di essi; generalmente i DB pubblicano un avviso sui loro siti web, all'interno di una lunga informativa sulla privacy, spiegando come accedere alle informazioni; tuttavia, questi avvisi possono essere difficili da comprendere. In risposta a una richiesta del consumatore, alcuni data broker forniranno il nome del consumatore accompagnato da alcune categorie di interesse generali associate a quel consumatore, come "appassionato di viaggi" o "amante del golf". Ai consumatori non viene fornito l'accesso a tutti i dati associati a loro e/o a tutte le inferenze fatte da quei dati. I data broker tipicamente forniscono l'accesso ai dati grezzi e non alle informazioni proprietarie che derivano attraverso inferenze. Di conseguenza, i consumatori potrebbero non sapere di essere stati classificati in un certo modo.

Nella misura in cui i consumatori possono accedere ai "propri" dati, devono fornire informazioni personali per verificare la loro identità e talvolta documentazione aggiuntiva tramite posta o e-mail, come una copia fisica o scansionata di un documento di identità con foto rilasciato dal governo o del passaporto.

Nello studio della FTC, i data broker affermano di utilizzare queste informazioni personali solo per scopi di autenticazione e per l'elaborazione delle richieste di accesso.

I data broker che vendono prodotti di marketing e forniscono ai consumatori l'accesso, offrono la possibilità di rinunciare all'uso o alla condivisione delle informazioni personali

dei soggetti per scopi di marketing. “Rinunciare” significa non visualizzare le informazioni personali del consumatore per i servizi di marketing del DB; potrebbero non cancellare le informazioni del consumatore dai loro sistemi per due differenti motivazioni:

- mantenere le informazioni per essere in grado di abbinare i record che potrebbero ricevere in futuro e identificare quali devono essere soppressi.
- continuare a utilizzare le informazioni sopresse in prodotti che visualizzano dati in forma aggregata e anonima, anche se, come affermato nella sezione 2 l’anonimizzazione sembra, comunque, essere aggirabile..

EXPERIAN

Experian per segmentare i propri consumatori utilizza un sistema denominato Mosaic. Esso è uno strumento di targhettizzazione di tipo statistico. Per quanto riguarda il mercato italiano, vengono classificati i consumatori in 13 tipologie e 41 gruppi in base a 700 variabili ottenute attraverso una pluralità di ricerche di mercato. È associato alle celle micro-censuarie ISTAT ma la predittività è basata sulle caratteristiche generali che sono presentate nella figura di seguito (EXPERIAN, 2024):

Dati demografici	Socio-economia	Posizione	Misure finanziarie	Caratteristiche dell'immobile
Età Stato civile Composizione familiare Durata della residenza Presenza di bambini Numero di occupanti Etnia Abilità linguistiche	Istruzione/qualifiche Occupazione Industria Ore lavorate Business casalingo Proprietà del veicolo	Urbanità/ruralità Mezzi di trasporto Tempo di viaggio per lavoro	Reddito Comportamento creditizio Proprietario di più case Previdenza/assistenza sociale	Mandato Valore dell'immobile Numero di stanze Anno di costruzione Numero di abitazioni Importo dell'affitto Quartieri di gruppo

FIGURA 20: Criteri di targhettizzazione per Experian

I gruppi e i tipi per i consumatori italiani sono:

Gruppo	Nome del gruppo	% Nuclei Familiari	Tipo	Nome dei Tipi	% Nuclei
A	Elite Urbane	8,74%	A01	Alta Borghesia	1,14%
			A02	Benestanti Smart	1,41%
			A03	Senior di Successo	2,96%
			A04	Professionisti Dinamici	3,23%
B	Benestanti di Provincia	7,19%	B05	Famiglie di Prestigio	1,42%
			B06	Ceto Medio Maturo	1,16%
			B07	Single Abienti Maturi	2,97%
			B08	Pensionati Attivi	1,64%

C	Giovani Urbani Emergenti	5,24%	C09	Rampanti Emancipati	1,93%
			C10	Giovani Famiglie Sprint	1,13%
			C11	Grandi Famiglie Moderne	2,18%
D	Ceto medio di Provincia	12,19%	D12	Giovani Famiglie Monoreddito	3,63%
			D13	Famiglie Affermate	3,31%
			D14	Imprenditori Tradizionali	5,25%
E	Giovani Urbani Svantaggiati	4,30%	E11	Famiglie in Aree Poco Popolate	2,39%
			E12	Famiglie in Aree Molto Popolate	1,91%
F	Pensionati Urbani in Affitto	5,64%	F17	Anziani di Terza Età	1,41%
			F18	Anziani di Quarta Età di Periferia	4,23%
G	Anziani di Provincia	5,73%	G19	Anziani Agiati	4,20%
			G20	Anziani Svantaggiati in Piccoli Paesi	1,08%
			G21	Senior precari	0,45%
H	Pendolari in Periferia	4,75%	H22	Famiglie in Periferia per Scelta	1,48%
			H23	Pendolari dell'industria	3,27%
I	Mediterranei Abienti	17,21%	I24	Pensionati al Mare	1,78%
			I25	Famiglie Mature Possidenti	5,44%
			I26	Giovani Famiglie Tradizionali	3,16%
			I27	Famiglie con Bassi Consumi	6,83%
J	Famiglie Rurali Svantaggiate	5,75%	J28	Lavoratori Precari	0,42%
			J29	Famiglie Rurali Digitali	1,47%
			J30	Nuclei Contadini del Sud	3,86%
K	Ceto Medio Rurale	8,08%	K31	Operai in età Pensionabile	2,15%
			K32	Famiglie in Aree Decentrate	2,28%
			K33	Famiglie Operaie	2,13%
			K34	Operai in Pensione	1,52%
L	Classe Rurale Moderna	8,32%	L35	Professionisti Autonomi di Successo	0,92%
			L36	Ceto Medio Montano	1,26%
			L37	Elite Intellettuali	1,15%
			L38	Ceto Medio di Provincia	4,99%
M	Rurali Tradizionali	6,86%	M39	Famiglie Rurali Semplici	0,48%
			M40	Anziani Rurali	4,46%
			M41	Anziani isolati	1,92%

Similmente, per il mercato statunitense, Experian divide i consumatori in 19 gruppi e 71 tipi come mostrato nella tabella successiva:

Gruppo	Nome gruppo	Codice tipo	Nome tipo
A	Elite "energetica"	A01	Nobiltà americana
		A02	Prosperità di platino
		A03	Bambini e Cabernet
		A04	Famiglie perfette
		A05	Coppie influenti
		A06	Urbanisti del jet set

B	Famiglie fiorenti	B07	Attraverso i secoli
		B08	Bambini e felicità
		B09	Famiglia "fun-tastic"
		B10	Successi cosmopoliti
C	Con fiducia	C11	Abitanti sofisticati delle città
		C12	"Golf cart" e buongustai
		C13	Filantropi sofisticati
		C14	"Boomers e Boomerang"
D	Suburbano	D15	Famiglie di "utilità sportiva"
		D16	Stabiliti in periferia
		D17	Diversità "cul de sac"
		D18	Vita notturna suburbana
E	Boomer fiorenti	E19	Consumatori esperti
		E20	"Nessun posto è come casa"
		E21	Splendore incontaminato
F	Famiglie promettenti	F22	Coppie in "corsia rapida"
		F23	Le famiglie contano di più
G	Giovani single	G24	Single ambiziosi
		G25	Margine urbano
H	Unione familiare borghese e multiculturale	H26	Assortimento progressivo
		H27	Vita di svago
		H28	Moderati di tutti i giorni
		H29	Divertimento in "destinazione"
I	Unione familiare borghese e multiculturale	I30	"Cene condivise e la natura"
		I31	Valori del duro lavoro
		I32	Confermitsti fermi
		I33	Equilibrio e armonia
J	"Anni d'autunno"	J34	Sofisticati suburbani
		J35	Fuga rurale
		J36	"Stabiliti e sensati"
K	Single "di spessore"	K37	Cablati per il successo
		K38	Mescolanza moderna
		K39	Fusione metropolitana
		K40	Ritmo bohemien
L	Boomer "del cielo blu"	L41	Consumistici e in espansione
		L42	Famiglie radicate nel "potere dei fiori"
		L43	Felicità casalinga
M	Famiglie	M44	Comfort creativo
		M45	Famiglie in crescita e in espansione
N	Orgoglio pastorale	N46	Famiglie con vera grinta americana
		N47	"Pragmatici rustici"
		N48	Beatitudine rurale del sud
		N49	Un tocco di tradizione
O	Single e neo-famiglie	O50	"A tutta velocità"
		O50	Abili nel digitale
		O52	Ambizione urbana
		O53	"Università e caffè"

		O54	Influenzati dagli influencer
		O55	"Soldati della famiglia"
P	"Connessioni Culturali"	P56	Medley di classe media
		P57	"Modesti metropolitanti medi"
		P58	Altire dell'eredità
		P59	Orizzonti in espansione
		P60	Avanzare con impegno
		P61	"Inizi semplici"
		Q	Guardiani degli anni d'oro
Q63	Senza vincoli e liberi in famiglia		
Q64	Stabiliti nella società		
Q65	Maturi e saggi		
R	"Fusione aspirazionale"	R66	Sognatori ambiziosi
		R67	Genitori appassionati
S	Con abitudini parsimoniose	S68	Sofisticati delle piccole città
		S69	Eredità urbane
		S70	Single parsimoniosi
		S71	Pensionati modesti

(Experian, 2019)

Experian fornisce anche le definizioni dei gruppi:

- "Power elite": le famiglie più ricche degli Stati Uniti, che vivono nei quartieri più esclusivi e si godono tutto ciò che la vita ha da offrire.
- Famiglie fiorenti: famiglie e coppie benestanti di mezza età che guadagnano redditi prosperi e hanno uno stile di vita molto confortevole e attivo.
- Con fiducia: coppie prospere e affermate nel loro periodo di massimo guadagno che vivono in case di periferia
- Suburbano: famiglie e coppie suburbane di mezza età, etnicamente miste, che guadagnano redditi elevati
- Boomer fiorenti: le coppie in età del "baby boom" della classe medio-alta che vivono stili di vita confortevoli e vivono in zone di periferia
- Famiglie promettenti: giovani coppie con figli, che vivono stili di vita incentrati sul bambino
- Giovani single: single più giovani e di mezza età che vivono stili di vita attivi ed energici nelle aree metropolitane
- Unione familiare borghese e multiculturale: coppie di mezza età e stabili economicamente che vivono in case di periferia
- "Crogiolo borghese": Reddito medio, famiglie di mezza età che vivono in case sostenute da solide occupazioni da operai
- "Anni d'autunno": coppie affermate e mature che vivono stili di vita gratificanti nelle case più vecchie
- Single "di spessore": single di età diversa che guadagnano redditi medi e sostengono stili di vita urbani attivi

- “Boomer del cielo blu”: famiglie della classe media in età da “baby boomer” che vivono in piccole città
- Famiglie in movimento: famiglie della classe operaia con figli piccoli, che guadagnano redditi moderati in comunità residenziali più piccole
- Orgoglio pastorale: mix eclettico di consumatori della classe medio-bassa che si sono stabiliti nelle aree di campagna e di piccole città
- Single e neo-famiglie: giovani single agli inizi della vita lavorativa e alcune famiglie costituite da poco, che vivono uno stile di vita cittadino
- “Connessioni Culturali”: famiglie diverse, a medio e basso reddito che vivono in appartamenti e residenze urbane
- Guardiani degli anni d'oro: pensionati che vivono in vecchie case, residenze stabili e comunità
- “Fusione aspirazionale”: single a basso reddito e genitori single che vivono in aree urbane e cercano di costruirsi una vita migliore
- Con abitudini parsimoniose: adulti attenti ai costi che vivono soli nelle aree urbane

Per ogni gruppo, inoltre, vengono fornite le seguenti informazioni dettagliate soprattutto di natura privata, riguardanti la targhettizzazione dei nuclei appartenenti ai vari gruppi, secondo il seguente schema:

- Valore della proprietà: stima del valore economico degli immobili in cui si vive.
- Reddito: livello di reddito medio.
- Livello di istruzione: grado di istruzione raggiunto e distribuzione.
- Come trascorrono il tempo libero in termini di attività preferite.
- Tipologie di investimenti finanziari più comuni.
- Tipologia e località dei viaggi più comuni (estero e/o all’interno degli States).
- Mezzi di informazione che utilizzano.
- Orientamento politico: tendenze politiche prevalenti e distribuzione delle preferenze politiche.
- “Come commercializzare con i consumatori”: quali canali utilizzare per raggiungere quel determinato gruppo/nucleo.

Di seguito sono riportate delle descrizioni fatte da Experian sui diversi gruppi:

“Per quanto riguarda le loro inclinazioni politiche, questi elettori attivi sono per lo più di centrodestra. Il 45% si considera repubblicano moderato e solo il 15% circa sono ultraconservatori di estrema destra”, “Poco meno del 75% delle famiglie possiede la propria casa, [...] le loro proprietà sono valutate solo alla metà della media nazionale. La maggior parte vive allo stesso indirizzo da meno di dieci anni, il che indica una tendenza allo spostamento in questo segmento”, “Sono più ricettivi agli annunci pubblicati sulla TV in

streaming, agli SMS mobili, alle e-mail e ai social media”, “Hanno abbastanza soldi dai loro investimenti regolari in azioni e obbligazioni di risparmio per concedersi una cena al ristorante”, “Questi acquirenti rappresentano un mercato forte per l'elettronica; acquistano tutti i tipi di apparecchiature video e audio”, “Quando escono a mangiare, è più probabile che si rechino in un ristorante locale o in una catena di fast food”, ...

Come analizzato nella sezione 1 queste sono le tipologie di dati più frequenti che i DB cercano, analizzano e vendono. Tuttavia, si trovano anche altre informazioni private, che non rientrano in quelle “tipiche” utilizzate dai DB, ad esempio:

“Molte case hanno un canestro da basket nel vialetto. Nei fine settimana, i marciapiedi sono pieni di adolescenti che fanno skateboard, vanno in bicicletta e tirano a canestro”, oppure “Con tassi bassi per l'uso delle carte di credito e degli sportelli bancomat, comunque, di solito pagano tutto in contanti”, oppure “Per l'eccitazione, tentano regolarmente la fortuna acquistando i biglietti della lotteria”.

3.5 I DATASET DEI DATA BROKER

I dataset utilizzati per le analisi di questo elaborato sono stati reperiti su AWS marketplace. In prima battuta, verranno analizzati singolarmente per comprendere come vengono offerti al pubblico e con che granularità vengono offerti. I primi ad essere analizzati sono quelli di Acxiom.

ACXIOM

1) Il dataset in questione: “Germany - Statistical Data for Marketing Areas (refined Postcode) SAMPLE” (AWS marketplace, 2024) è composto da un sample di variabili (senza record specifici); sono “informazioni per un'ampia gamma di comportamenti. Aiuta a comprendere i potenziali di mercato e a gestire le posizioni dei siti, nonché a indirizzare le attività di marketing”.

- **Campi:**
 - Popolazione per genere
 - Popolazione per 17 fasce di età
 - Popolazione per stato civile
 - Popolazione per livello di istruzione
 - Estensione della superficie in km²
 - Famiglie in totale
 - Famiglie per dimensione

- Famiglie per presenza di figli
- Famiglie per stato sociale
- Famiglie per fasce di reddito
- Potere d'acquisto totale nell'area e potere d'acquisto pro capite
- Previsioni sull'età della popolazione negli anni
- Numero totale di edifici
- Edifici per uso abitativo, commerciale o misto
- Edificio per numero di unità abitative
- Numero totale di auto
- Autovetture ad uso privato o commerciale
- Popolazione diurna
- Frequenza classificata dei passanti

Inoltre, vi è la presenza di variabili molto specifiche sulle affinità del nucleo preso in considerazione:

Categoria	Sottocategoria	Codice	Descrizione
<i>sociodemografia: propensioni</i>	Affinità	SCO BAWE	Affinità alla bozza di birra
<i>sociodemografia: propensioni</i>	Affinità	SCO CHOC	Affinità al cioccolato
<i>sociodemografia: propensioni</i>	Affinità	SCO ESPR	Affinità all'espresso
<i>sociodemografia: propensioni</i>	Affinità	SCO GART	Affinità al giardinaggio
<i>sociodemografia: propensioni</i>	Affinità	SCO OEKO	Affinità alla green energy
<i>sociodemografia: propensioni</i>	Affinità	SCO ONLI	Affinità alla "connessione online"
<i>sociodemografia: propensioni</i>	Affinità	SCO PATV	Affinità alla "pay TV"
<i>sociodemografia: propensioni</i>	Affinità	SCO PHON	Affinità agli smartphone
<i>sociodemografia: propensioni</i>	Affinità	SCO VERA	Affinità agli ordini per corrispondenza
<i>sociodemografia: propensioni</i>	Affinità	SCO ZURV	Affinità all'assicurazione integrativa privata per pensione

▪ Granularità

I dati statistici anonimi sono suddivisi nelle aree di marketing standard di Acxiom per la Germania, con almeno 4 nuclei familiari per ciascuna area geografica. Il dataset contiene dati relativi a 39 delle oltre 100.000 aree di marketing. Il livello dell'area di marketing combina il codice postale tedesco con le aree statistiche ufficiali utilizzate dalle comunità per la pianificazione interna. Questa combinazione, se necessario, viene ulteriormente suddivisa per creare aree omogenee con una media di 400 nuclei familiari. Mentre i codici postali tedeschi coprono circa 8000 aree con una media di poco più di 10000 persone ciascuna, le aree di marketing comprendono oltre 100.000 aree con una media di circa 700

persone. L'utilizzo del codice postale facilita l'inclusione dei dati a quel livello nell'analisi, mentre le aree statistiche ufficiali rendono semplice il confronto dei dati di mercato con altre statistiche per la Germania. Di conseguenza, Acxiom non si limita ad offrire i propri dati ad un'unica granularità. Di seguito sono riportati i livelli di granularità del dataset. Sono stati suddivisi in 6 gruppi per analizzare più agevolmente i dati:

Gruppo	A	B	C	D	E	F
Numero di aree marketing	>24.000.000	>1.500.000	> 110.000	>35.000	>11.000	>8.100
Numero di persone per area	3,5	~55	~700	~2500	~7500	~10000

Alcune variabili possono essere presenti con ogni precisione, ad esempio, la variabile "federal state" è presente per tutte le aree di marketing a qualsiasi livello di specificità; oppure solamente un valore: "B1" che rappresenta la popolazione di età pari a un anno, ha come granularità la D; oppure in più di un valore: "classified affinity to chocolate" ha come granularità D ed E. Da notare che il gruppo A dà informazioni sulla singola proprietà (appartamento, casa indipendente, ufficio, ...).

- **Abbonamento**

L'abbonamento prevede un mese di iscrizione gratuita per i dictionary, successivamente se si vuole comprare il dataset reale bisogna comprarlo per usufruirne.

- **Frequenza di aggiornamento**

I dati vengono aggiornati su base annuale e gli aggiornamenti non sono inclusi nell'offerta gratuita. Si legge: "se sei interessato agli aggiornamenti automatici, visita Acxiom o contattaci".

Come affermato in precedenza, i DB consigliano la consulenza individuale per vendere i propri servizi.

2) Un altro dataset è: "Germany Statistical Data for Municipalities SAMPLE" (AWS marketplace, 2024). Esso riguarda l'estensione delle aree tedesche, la densità di popolazione, la valutazione del patrimonio immobiliare, nonché può essere utile per comprendere il comportamento dei consumatori o pianificare le posizioni dei siti.

- **Granularità**

I dati statistici anonimi basati su fonti ufficiali vengono visualizzati a livello dei comuni tedeschi (Gemeinden). Il file contiene dati per 4 di tutti gli oltre 11.000 comuni tedeschi con una media di circa 7.5000 persone.

- **Campi**

Il file include i seguenti campi:

- Codice per il comune

- Nome del comune
- Estensione della superficie in km²

EQUIFAX

Equifax invece, offre l'accesso a diversi "Analytic Dataset" (AD) nei quali si possono trovare tipologie di dataset (sample) diverse in base ai diversi settori. Inoltre, in molti dataset che analizzano gli stessi servizi a granularità differenti, Equifax offre una chiave univoca che permette di legare più dataset: "integra con altri set di dati Equifax disponibili su questo marketplace per ottenere visualizzazioni aggiuntive."

1) Nel primo AD (AWS marketplace, 2024) si possono trovare "dati granulari a livello di consumatore, anonimi e non aggregati, per tutte le classi di asset; esso fornisce dati a livello di prestito attraverso cicli di credito e classi di attività, tra cui auto, carte di credito, mutui e prestiti studenteschi per prevedere con maggiore precisione le prestazioni future".

Esso è creato da un campione statistico imparziale del 10% della popolazione attiva nel credito statunitense.

Questo AD include un identificatore anonimo del consumatore che consente agli acquirenti di collegare il comportamento degli stessi consumatori nel tempo. Il set di dati analitici consente di tenere traccia dei cambiamenti nei codici postali degli indirizzi dei consumatori nel tempo e include anche un numero di conto codificato che consente agli utenti di collegare la performance di ciascun prestito nel tempo.

- **Campi inclusi**
 - Punteggio del rischio di credito
 - Fascia d'età
 - Geografia
 - Tipo di account
 - Saldo del debito
 - Stato di delinquenza
 - Punteggio del reddito dei consumatori
 - Indagini

- **Granularità**

Il dataset fornisce dati granulari a livello di prestito su cicli di credito e classi di attività, tra cui prestiti auto, carte di credito, mutui, prestiti agli studenti e prestiti personali non garantiti per prevedere con maggiore accuratezza le performance future. È creato da un

campione statistico imparziale del dieci per cento della popolazione attiva di credito degli Stati Uniti su tutti i confini geografici, con dati storici a partire dal 2005.

2) Il secondo AD è "Credit Trends Free Dataset" (AWS marketplace, 2024) è un campione di 100 righe che rappresenta il dataset completo in termini di tipi di dati, organizzazione e campi.

Credit Trends è un set di dati anonimizzato proveniente dal database di credito al consumo Equifax. "Evidenzia le transazioni di prestiti al consumo segnalate a Equifax da banche, unioni di credito e altre società finanziarie. Sfrutta questi dati per comprendere le modifiche ai conti e ai saldi, nonché le tendenze di insolvenza, l'utilizzo da parte dei consumatori e i tassi di svalutazione per un settore.

I set di dati possono essere valutati e condivisi come singoli elementi, per coloro che desiderano solo un sottoinsieme o un file completo Credit Trends."

▪ **Campi inclusi:**

- Data di origine e anno di origine
- Saldo di origine
- Origine e punteggio di rischio attuale
- Numero e dollari delle transazioni 30, 60, 90, 120, chiuse, fallimentari, ecc.
- Importo del pagamento programmato
- Alto merito creditizio

Equifax suggerisce di utilizzare il AD per i seguenti casi d'uso:

▪ **Analisi dell'impatto economico**

Comprendere e identificare gli indicatori principali della salute economica.

▪ **Analisi quantitativa**

Misurare direttamente le tendenze del mercato e identificare i punti di inflessione.

▪ **Analisi del rischio**

Stabilire metriche/linee di base personalizzate per inadempienza e morosità.

Monitorare le prestazioni in base alla categoria di prestito, punteggio di rischio, annata, area geografica.

▪ **Segmentazione del pubblico**

Sovrapporre i dati demografici dei consumatori, come età e reddito, per comprendere meglio il comportamento del credito al consumo in base al segmento.

3) Il terzo AD è denominato: “Real Property Insights” (AWS marketplace, 2024). “Per capitalizzare l'investimento e massimizzare le entrate, i proprietari di case/appartamenti e gli investitori hanno bisogno di informazioni più dettagliate sullo stato finanziario dei loro inquilini e di proprietà simili nelle aree circostanti. Questo set di dati demo contiene oltre 100 record di dati fittizi estremamente simili ai reali”.

AMP Insights di Equifax è una soluzione confezionata che offre uno "sguardo all'interno delle mura" di appartamenti e proprietà multifamiliari. Riunisce dati su inquilini, proprietà e mercato in un'unica soluzione per un'analisi più semplice

Equifax suggerisce l'utilizzo di questo AD per valutare il valore degli investimenti immobiliari commerciali, identificare il potenziale di mercato e le opportunità di crescita, mitigare il rischio e realizzare più rapidamente il ritorno sugli investimenti.

Il set di dati AMP Insights fornisce informazioni aggregate tra cui:

Affittuario/Conduttore (presentato come media per l'appartamento):

- Punteggio di credito
- Reddito stimato dell'inquilino
- Potere di spesa
- Indice di ricchezza
- Durabilità finanziaria
- Dimensione della famiglia
- Età del capofamiglia
- Famiglie con bambini
- Nome e indirizzo della proprietà
- Numero di edifici e numero di unità
- Dimensioni della proprietà e metratura dell'edificio
- Tipo di posizione
- Proprietario dell'edificio
- Valore di mercato stimato dell'immobile
- Reddito familiare
- Numero di famiglie
- Popolazione

- Le attività commerciali della zona

4) Il quarto set di dati Real Property Insights (AWS marketplace, 2024) fornisce una panoramica completa dei dati di pubblico dominio sugli immobili residenziali negli Stati Uniti, tra cui informazioni sulla proprietà, valori e caratteristiche degli immobili, date e prezzi delle vendite, privilegi ipotecari, posizioni azionarie e inserzioni attive.

Queste informazioni possono essere utilizzate per benchmarking, ricerca e analisi, trending, previsioni, marketing.

▪ **Campi inclusi:**

- Bandiera di occupazione del proprietario
- Valore attuale della casa
- Caratteristiche della casa come tipo di proprietà, numero di camere da letto e bagni, metratura della casa e del lotto e altro ancora
- Ultima data di vendita e prezzo per la durata della residenza e identificazione del nuovo proprietario
- Privilegi ipotecari attivi, inclusi tassi di interesse, tipo di prestito e nome del prestatore
- Dati di quotazione attivi come prezzo, data e giorni sul mercato

5) Il quinto ed ultimo AD è "Housing Characteristics" nel quale si trovano caratteristiche abitative dagli uffici del County Assessor su proprietà residenziali (fino a 4 unità) negli Stati Uniti.

Equifax offre la possibilità di essere contattata per ricevere richieste di diversa geografia (ad esempio stato, contea, MSA- Metropolitan Statistical Area) e frequenza di aggiornamento (ad esempio mensile, trimestrale, annuale).

Inoltre, se l'acquirente vuole informazioni più approfondite, Equifax propone il pacchetto formato da Property Listings, Recent Home Purchase Transactions, Home Value Trends and Forecasts e Home Equity & Mortgages.

▪ **Campi inclusi:**

- Tipologia di immobile (ad esempio monofamiliare, condominio)
- Conteggio delle camere da letto e dei bagni
- Dimensioni della casa e metratura del lotto

- Anno di costruzione

EXPERIAN:

Experian WorldView (AWS marketplace, WorldView Experian, 2024) fornisce attributi demografici a livello geografico per griglie di 250 x 250 metri che coprono il globo. È consigliato agli utenti business per prendere decisioni informate sui mercati internazionali, accedendo a una serie di pacchetti di dati coerenti tra diverse nazioni. Dati sociodemografici aggregati, spesa dei consumatori e punti di interesse per posizioni geografiche in oltre 90 paesi.

I dati contengono (tra gli altri attributi):

- Codice Paese
- Codice di Rete
- Area Amministrativa
- Popolazione Totale
- Numero di Famiglie
- Popolazione Maschile
- Popolazione Femminile
- Popolazione per Età (fasce di 15 anni)
- Popolazione per Età e Genere (fasce di 15 anni)
- Popolazione Disoccupata
- Reddito Disponibile
- Reddito Disponibile Pro Capite

ENIGMA

Il dataset offerto da Enigma: “NYC Property Land Use Data (Pluto) - As of Sept 2019” (AWS marketplace, 2024) offre oltre 70 attributi su ciascuno degli oltre 850.000 lotti immobiliari di New York, tra cui la metratura, il nome del proprietario, l'anno di costruzione e il valore fiscale stimato. Questo dataset è basato su informazioni di pubblico dominio, sono dati offerti dal dipartimento delle finanze e lo strumento è gestito dal dipartimento della pianificazione della città di New York.

Campi chiave

Codice	Descrizione
borough	Distretto della città di New York
block	Numero del blocco nella mappa catastale
lot	Numero del lotto all'interno del blocco
cd	Distretto comunitario
bct2020	Blocco censuario del 2020
zipcode	Codice postale
address	Indirizzo dell'immobile

splitzone	Zona divisa
bldgclass	Classe dell'edificio
landuse	Uso del suolo
ownertype	Tipo di proprietario
ownername	Nome del proprietario
lotarea	Superficie del lotto
bldgarea	Superficie dell'edificio
comarea	Superficie commerciale
resarea	Superficie residenziale
officearea	Superficie uffici
retailarea	Superficie vendita al dettaglio
garagearea	Superficie garage
strgearea	Superficie magazzino
factoryarea	Superficie fabbrica
numbldgs	Numero di edifici
numfloors	Numero di piani
unitsres	Numero di unità residenziali
unitstotal	Numero totale di unità
lotfront	Fronte del lotto
lotdepth	Profondità del lotto
bldgfront	Fronte dell'edificio
bldgdepth	Profondità dell'edificio
ext	Estensione
assesland	Valore catastale del terreno
assesstot	Valore catastale totale
yearbuilt	Anno di costruzione
borocode	Codice del distretto
bbl	Numero del blocco e lotto
condono	Numero del condominio
tract2010	Tratto censuario del 2010
xcoord	Coordinata X
ycoord	Coordinata Y
taxmap	Mappa catastale
latitude	Latitudine
longitude	Longitudine

3.6 GEOGRAFIA DEI DATASET

I DB che utilizzano la piattaforma AWS marketplace vendono i propri servizi per più aree geografiche. Ad esempio, Acxiom offre il prodotto “consumer segmentation” per Germania, Svezia, Singapore, Ungheria, Australia, India, Irlanda, Giappone, Spagna, Sud Africa, Belgio, Hong Kong, Italia, Brasile, Svizzera, Nuova Zelanda, Canada, Finlandia, Slovacchia, Repubblica Ceca, Francia, Norvegia, Portogallo, Danimarca, Messico, ... (AWS marketplace, Acxiom consumer segmentation, 2024)

Experian offre “WorldView – Demographics” per Perú, Algeria, Philippine, South Africa, Iran, Chile, India, Colombia, Indonesia, Egitto, Germania, Nigeria, Kazakhstan, Turchia, Arabia Saudita, Messico, Argentina, Ucraina, Marocco, Pakistan, ... (AWS marketplace, WorldView Experian, 2024).

Ogni DB ha vari fornitori in molte nazioni, basti pensare che i principali player sono attivi su almeno quattro continenti. Non ci sono, dunque, evidenze che dimostrino la presenza di area geografiche in cui non vi sia competizione tra DB.

Per quanto riguarda la granularità dei dati offerti dai vari DB, bisogna confrontarli singolarmente e comprendere se si presenti la possibilità di legarli. Come supposto nei modelli analizzati nella sezione 1. Di seguito l’analisi dei dataset.

3.7 SOVRAPPOSIZIONI E CONFRONTO TRA DATASET

Per il confronto tra i dataset sono stati presi in considerazione quelli di Acxiom, Equifax ed Enigma.

Premesse e assunzioni

1. Il dataset di Acxiom offerto gratuitamente è riferito alla popolazione tedesca; tuttavia, la struttura del dataset dovrebbe essere pressoché la stessa anche per il mercato statunitense (nello specifico di New York).
2. Sono stati utilizzati più dataset di Equifax, che sono stati legati da un attributo che assume il valore di chiave, in particolare: “BLS Property ID”. Esso è un identificatore univoco della singola abitazione (casa indipendente, appartamento, ufficio, ...) e viene utilizzato dal DB per permettere ai clienti di effettuare analisi orizzontali sui diversi dataset.
3. I dataset di Equifax sono in riferimento al mercato delle proprietà statunitensi, di conseguenza, rientra anche la città di New York.
4. Il dataset di Enigma, che lo Stato di New York offre anche indipendentemente e gratuitamente, permette di avere informazioni preziose sui proprietari delle abitazioni; tuttavia, non verrà preso in considerazione quando si farà il confronto tra i dataset dei DB.
5. Negli Stati Uniti, i dati catastali, nello specifico il nome dei proprietari delle abitazioni, sono pubblici e non vi è alcuna restrizione; in Italia, e in generale all’interno dell’UE, alcuni dati catastali non sono pubblici; tuttavia, se si dimostra un interesse legittimo possono essere richiesti ed ottenuti. Visto che molti DB hanno come partner o clienti agenzie immobiliari, è logico pensare che riescano ad ottenere questi dati anche se non hanno il diritto di venderli.
6. Durante le analisi successive bisogna tenere in considerazione che i dataset di Enigma ed Equifax hanno una granularità univoca (e comune), ovvero quella di singolo edificio/abitazione; ciò significa che ogni variabile è riferita ad

esattamente una singola abitazione; invece, le variabili di Acxiom possono essere disponibili in più granularità.

SOVRAPPOSIZIONI

Le variabili sono state suddivise in macroaree semantiche e per ognuna di esse sono state identificate le variabili perfettamente sovrapponibili. Le celle dello stesso colore sono relative alle variabili comuni tra i vari dataset.

- **Geolocalizzazione**

ACXIOM		ENIGMA		EQUIFAX	
Codice	Descrizione	Codice	Descrizione	Nome variabile	Descrizione
KGS02	Stato federale	borough	Distretto della città di New York	Corrected Address Full Address	Indirizzo completo
KGS05	Distretto	block	Numero del blocco nella mappa catastale	Corrected Address House Number	Numero civico
KGS08	Città	lot	Numero del lotto all'interno del blocco	Corrected Address Street Direction	Direzione della strada
KGS12	Parti della città	cd	Distretto comunitari o	Corrected Address Street Name	Nome della strada
KGS14PLZ	Area di marketing KGS14PLZ	bct2020	Blocco censuario del 2020	Corrected Address Street Type	Tipo di strada
KGS16	Segmenti di strada	zipcode	Codice postale	Corrected Address Street Post Direction	Direzione postale della strada
GEMEINDE	Comune	address	Indirizzo dell'immobile	Corrected Address Unit Type	Tipo di unità
MIKROMARKT	Micro mercato	splitzone	Zona divisa	Corrected Address Unit Number	Numero di unità
NAME	Nome dell'area di marketing	borocode	Codice del distretto	Corrected Address City	Città
ORT	Insedimento	latitude	Latitudine	Corrected Address State	Stato
ORTSTEIL	Parte dell'insediamento, numero postale corretto	longitude	Longitudine	Corrected Address Zip	Codice postale
PLZ	Codice postale			Subject Latitude	Latitudine della proprietà

STAB	Micro mercato			Subject Longitude	Longitudine e della proprietà
STRASSE	Strada			AREA_NUMBER_OF_BUSINESSES_IN_ZIP	Numero di attività commerciali nel codice postale
ALLORT	Codifica del comune da parte della posta tedesca			AREA_NUMBER_OF_PEOPLE_IN_ZIP	Numero di persone nel codice postale
HAUSNR	Numero civico			AREA_NUMBER_OF_HH_IN_ZIP	Numero di famiglie nel codice postale
				AREA_AVG_HH_INCOME_IN_ZIP	Reddito medio delle famiglie nel codice postale

▪ **Informazioni di varia natura sull'edificio**

ENIGMA		EQUIFAX		ACXIOM	
Codice	Descrizione	Nome variabile	Descrizione	Codice	Descrizione
bldgclass	Classe dell'edificio	BLS_ID	Identificatore	GEBÄUDE	Numero di edifici
landuse	Uso del suolo	Owner Occupancy Flag	Indice di occupazione	GEB_GEW	Edificio ad uso commerciale
lotarea	Superficie del lotto	Property Type	Tipo di proprietà	GEB_KA	Edificio senza informazioni sull'uso
bldgarea	Superficie dell'edificio	Year of Construction	Anno di costruzione	GEB_MIX	Edifici a uso misto
comarea	Superficie commerciale	Acres	Acri	GEB_PRIV	Edificio ad uso privato
resarea	Superficie residenziale	Number of Garage Spaces	Numero di posti auto nel garage	PHU_AVGA	Area media degli edifici principali
officearea	Superficie uffici	Garage Square Footage	Superficie del garage in piedi quadrati	PHU_DENS	Densità degli edifici
retailarea	Superficie vendita al dettaglio	Main Square Footage	Superficie principale in piedi quadrati	PHU_SUMA	Somma totale di tutte le aree degli edifici

garagearea	Superficie garage	Basement Finished Square Footage	Superficie del seminterrato finito in piedi quadrati	MIETSCORE	Probabilità che un edificio sia affittato
strgearea	Superficie magazzino	Basement Square Footage	Superficie del seminterrato in piedi quadrati	WO_EIG	Appartamenti di proprietà (totale) 2015
factryarea	Superficie fabbrica	Subdivision Name	Nome della suddivisione	WO_MIE	Appartamenti in affitto (totale) 2015
numbldgs	Numero di edifici	Number of Stories	Numero di piani		Appartamenti con n stanze 2016
numfloors	Numero di piani	Number of Bathrooms	Numero di bagni	IND	Numero medio di appartamenti per edificio
unitsres	Numero di unità residenziali	Number of Bedrooms	Numero di camere da letto	WG_TO	Edifici residenziali (totale)
unitstotal	Numero totale di unità	Number of Fireplaces	Numero di camini		
lotfront	Fronte del lotto	Total Rooms	Numero totale di stanze		
lotdepth	Profondità del lotto	Property Style	Stile della proprietà		
bldgfront	Fronte dell'edificio	PROPERTY_MSA	Area statistica metropolitana		
bldgdepth	Profondità dell'edificio	PROPERTY_TYPE_LAND_USE_CODE	Codice di utilizzo del terreno		
ext	Estensione	Standardized Land Use Code	Codice standardizzato di uso del terreno		
yearbuilt	Anno di costruzione	property_number_of_hh_w_kids	Numero di famiglie con bambini		
condono	Numero del condominio	property_avg_hh_occupancy_in_months	Occupazione media delle famiglie in mesi		

- **Valore dell'edificio**

Acxiom non offre alcun dato sui valori dell'edificio.

ENIGMA		EQUIFAX	
Codice	Descrizione	Nome variabile	Descrizione
assesland	Valore catastale del terreno	Assessed Value (total)	Valore totale stimato
assesstot	Valore catastale totale	Assessment Year	Anno di valutazione
		Assessed Improvement Value	Valore stimato delle migliorie
		Assessed Land Value	Valore stimato del terreno
		Estimated LTV	Loan-to-Value stimato
		Estimated CLTV	Combined Loan-to-Value stimato
		Last Sale Price	Prezzo dell'ultima vendita
		Last Sale Date	Data dell'ultima vendita
		Listing Status	Stato dell'inserzione
		Days On Market	Giorni sul mercato
		Listing Price	Prezzo di vendita

▪ **Situazione finanziaria**

EQUIFAX		ACXIOM	
Nome variabile	Descrizione	Codice	Descrizione
1st Lien Amount	Importo del primo diritto di pegno	HH_U900	Famiglie con reddito netto mensile < 900€
1st Lien Date	Data del primo diritto di pegno	HH_U1300	reddito netto mensile da 900€ a meno di 1300€
1st Lien Interest Rate	Tasso di interesse del primo diritto di pegno	HH_U1500	reddito netto mensile da 1300€ a meno di 1500€
1st Lien Rate Type	Tipo di tasso del primo diritto di pegno	HH_U2000	reddito netto mensile da 1500€ a meno di 2000€
1st Lien Type	Tipo di primo diritto di pegno	HH_U2600	reddito netto mensile da 2000€ a meno di 2600€
Estimated 1st Lien Balance	Saldo stimato del primo diritto di pegno	HH_U3600	reddito netto mensile da 2600€ a meno di 3600€
2nd Lien Amount	Importo del secondo diritto di pegno	HH_U5000	reddito netto mensile da 3600€ a meno di 5000€
2nd Lien Date	Data del secondo diritto di pegno	HH_A5000	reddito netto mensile superiore a 5000€
2nd Lien Interest Rate	Tasso di interesse del secondo diritto di pegno		
2nd Lien Rate Type	Tipo di tasso del secondo diritto di pegno		
2nd Lien Type	Tipo di secondo diritto di pegno		
Estimated 2nd Lien Balance	Saldo stimato del secondo diritto di pegno	EHK_IDX	Indice del potere d'acquisto speso per commercio al dettaglio
Property_avg_tenants_score	Punteggio medio degli inquilini	V_AKTIE	Azioni (valore di mercato) in €

Property_pct_of_tenants_score_below_650	Percentuale di inquilini con punteggio inferiore a 650	V_AKTIVA	Prestiti (attivi) in €
Property_pct_of_tenants_score_650_700	Percentuale di inquilini con punteggio tra 650 e 700	V_BAU_S	Risparmi in €
Property_pct_of_tenants_score_above_700	Percentuale di inquilini con punteggio superiore a 700	V_GVM	Spese per accumulare risorse monetarie in €
Property_avg_households_income	Reddito medio delle famiglie	V_HYPO	Mutui ipotecari in €
Property_pct_of_hh_income_below_52k	Percentuale di famiglie con reddito inferiore a 52.000 USD	V_INV	Certificato di fondi di investimento (valore di mercato) in €
Property_pct_of_hh_income_52k_156k	Percentuale di famiglie con reddito tra 52.000 USD e 156.000 USD	V_KNSMK	Credito al consumo in €
Property_pct_of_hh_income_above_156k	Percentuale di famiglie con reddito superiore a 156.000 USD	V_PASSIV	Depositi (passivi) in €
		V_KFZ	Assicurazione auto in €
		V_LV	Assicurazione sulla vita in €

▪ Socio-demografia

Per quanto riguarda i dati sociodemografici, Equifax dà informazioni solo sull'età media degli inquilini dell'edificio. Acxiom, d'altra parte, ha variabili che coprono quasi interamente le possibili combinazioni di genere ed età visto che presenta i dati in diversa granularità; ad esempio: popolazione totale, popolazione maschile suddivisa per fasce d'età, popolazione femminile suddivisa per fasce d'età, popolazione per fasce d'età: gruppi specifici (es. 0-3 anni, 3-6 anni, ecc.), popolazione under 18, popolazione over 18, popolazione over 65, età media, indice di popolazione su densità e sviluppo, popolazione divorziata, popolazione single, popolazione di non sposati, "numero di separate, risposate", popolazione vedova, popolazione in coppie sposate, livelli di istruzione, numero di iscritti alle scuole superiori e università, stranieri totali, popolazione straniera maschile e femminile, densità di popolazione in km^2 , nascite totali per anno, decessi totali per anno, numero di persone che lasciano la regione, numero di persone che arrivano nella regione, dati storici di popolazione, proiezioni di popolazione: stime della popolazione per anni futuri (es. 2025, 2030, ecc).

L'unica sovrapposizione è, dunque:

ACXIOM		EQUIFAX	
Codice	Descrizione	Nome variabile	Descrizione
MTALTS	Età media	PROPERTY_AVG_TENANTS_AGE	Età media degli inquilini

Per entrambi i dataset la granularità delle due variabili è a livello di singola abitazione.

Le variabili appartenenti alle macroaree di cui sopra, sono tutte quelle per le quali vi è sovrapposizione.

CONFRONTO

Si vuole dimostrare la complementarità, anche parziale, dei dataset dei DB. Tenendo in considerazione che tali analisi sono state effettuate su dataset sample, si possono, in ogni modo, ottenere risultati che siano incipit di spunti di riflessioni interessanti.

Ad esempio, si effettui un'operazione di "join" tra i dataset di Acxion e Equifax e si utilizzi come chiave, ad esempio: "street", "lat, long", "district". Tramite una di queste chiavi i dataset saranno legati. Si supponga adesso di effettuare il "join" tramite la chiave "lat, long"

Dataset di Acxiom

Per quanto riguarda Acxiom, le variabili considerate sono:

Codice	Descrizione
X	X-coordinate (WGS84)
Y	Y-coordinate (WGS84)
KGS05	district
SCO OEKO	Affinità classificata all'energia verde (oekostrom)
HH_U900	Nuclei familiari con reddito netto mensile inferiore a 900€
HH_U1300	Nuclei familiari con reddito netto mensile da 900€ a meno di 1300€
HH_U1500	Nuclei familiari con reddito netto mensile da 1300€ a meno di 1500€
HH_U2000	Nuclei familiari con reddito netto mensile da 1500€ a meno di 2000€
HH_U2600	Nuclei familiari con reddito netto mensile da 2000€ a meno di 2600€
HH_U3600	Nuclei familiari con reddito netto mensile da 2600€ a meno di 3600€
HH_U5000	Nuclei familiari con reddito netto mensile da 3600€ a meno di 5000€
HH_A5000	Nuclei familiari con reddito netto mensile superiore a 5000€

Dataset di Equifax

Per Equifax, invece:

Nome variabile	Descrizione
Subject Latitude	Latitudine della proprietà
Subject Longitude	Longitudine della proprietà
Assessed Value (total)	Valore totale stimato
Assessed Land Value	Valore stimato del terreno
Last Sale Price	Prezzo dell'ultima vendita
Last Sale Date	Data dell'ultima vendita
Listing Price	Prezzo di vendita

Main Square Footage	Superficie principale in piedi quadrati
Number of Stories	Numero di piani
Number of Bathrooms	Numero di bagni
Number of Bedrooms	Numero di camere da letto
Number of Fireplaces	Numero di camini
Total Rooms	Numero totale di stanze
Property Style	Stile della proprietà

Avendo adesso un unico dataset, si può effettuare, ad esempio, un'analisi relativa alla vendita di impianti fotovoltaici. L'analisi congiunta del reddito mensile per abitazione (Axiom) e degli indicatori sul valore dell'immobile (Equifax), può essere indicativa del livello di ricchezza. Di questa analisi, si può ampliare la granularità effettuando una "group by" per district. Intersecando i distretti "ricchi" con quelli che presentano un'alta propensione per la green energy, si possono ricavare le informazioni sulla struttura delle abitazioni appartenenti a suddetti distretti dal dataset di Equifax e l'azienda in questione sarebbe, così, in grado di offrire preventivi di fotovoltaici specifici (in termini di dimensioni).

Da qui si deduce la necessità da parte dell'impresa di acquisire entrambi i dataset per effettuare analisi così dettagliate. Inoltre, si deduce intuitivamente, come l'elevata precisione di informazioni che l'azienda possiede, consenta alla stessa di addebitare il prezzo che ne massimizza il profitto e l'estrazione del surplus del consumatore.

Per quanto riguarda, invece, i dataset di Equifax ed Enigma, possono essere effettuate ulteriori analisi. I due dataset presentano, come attributi in comune, latitudine e longitudine, oltre a tutti gli identificatori per la singola abitazione (via, numero civico, ...).

Dunque, considerando le due coordinate geografiche come chiavi, si può effettuare un'operazione di "join" tra i due. Nel dataset offerto da Enigma, essendo dati catastali, è presente il nome del/della proprietario/a dell'edificio in analisi.

Di conseguenza, a catena, si può risalire ai nomi di tutti i proprietari di un determinato distretto, e, supponendo che siano anche inquilini, per quell'insieme di proprietari, si potrebbe anche conoscere se, in quel distretto, vi è un'alta affinità al cioccolato, al giardinaggio o alla birra alla spina.

3.8 LA RISPOSTA DEL MERCATO AI DB

Per contrastare l'invasione della privacy, sono nate aziende specializzate nel proteggere i dati online. Queste aziende offrono servizi per eliminare o ridurre la presenza digitale di individui, limitando l'accesso dei data broker alle informazioni e permettendo ai propri clienti di contrastare l'elevata capacità di estrazione del surplus.



FIGURA 21: Logo Incogni (Incogni, 2024)

Un esempio di tali aziende è Incogni. Incogni offre un servizio in abbonamento che aiuta gli utenti a esercitare il proprio diritto di imporre ai DB di eliminare le proprie informazioni personali.

Incogni rimuove automaticamente i dati degli utenti da centinaia di database di data broker ed effettua controlli regolari per garantire che questi ultimi non aggiungano nuovamente le informazioni.

La mission di Incogni è “aiutarti a riprendere il controllo della privacy dei tuoi dati!”. Svolge le seguenti azioni: crea un elenco di DB che potrebbero disporre delle informazioni di ciascun cliente e invia richieste di rimozione a tali società. Inoltre, dà seguito alle richieste respinte, collaborando anche con le agenzie di tutela dei consumatori, e ripete regolarmente il processo di rinuncia per garantire che i dati del cliente rimangano fuori dal mercato.

Ci si potrebbe chiedere se il business di Incogni sia legittimo. Incogni è un'entità legale legittima che segue le linee guida delle regolamentazioni sulla privacy dei dati come il California Consumer Privacy Act (CCPA), il Regolamento generale sulla protezione dei dati (GDPR) e il GDPR del Regno Unito.

In conclusione, Incogni tratta con i data broker per conto dell'utente in modo che non debba essere in prima persona a trovare e richiedere continuamente ai DB di eliminare i propri dati. (Incogni, 2024)

Databroker.global



FIGURA 22: Logo Databroker.global (Databroker.global, 2024)

Un'altra piattaforma nata di recente (la versione commerciale è stata rilasciata nel 2020) è “databroker.global” che permette a qualsiasi entità di vendere i propri dati (anche ad un prezzo pari a 0). Si tratta di un vero e proprio mercato dei dati accessibile anche ai consumatori che possono comprare e vendere dati, domandati e offerti con la piena

consapevolezza. Dalla sezione “documentation” del sito si evince che alcuni player dell'economia dei dati sono ricchi di dati, mentre altri ne sono poveri. La capacità di scambiarli efficacemente indipendentemente dalle dimensioni potrebbe colmare il divario tra i partecipanti ricchi di dati e quelli poveri. Così facendo, si migliora e si espande l'applicazione dei dati, aumentando il valore e l'utilità di questo asset per tutti. Pertanto, mercati robusti per i dati possono essere considerati infrastrutture essenziali per l'economia dei dati, che è in rapida crescita. Non per caso, infatti, è stato scelto “databroker” come nome della piattaforma.

Databroker è “un mercato decentralizzato supportato da blockchain per i dati in cui acquirenti e venditori scambiano dati su base peer-to-peer”. L'obiettivo dichiarato è quello di creare un mercato aperto, sicuro, trasparente ed efficiente per i dati.

La struttura di pricing suggerita dalla piattaforma riprende le strategie utilizzate dai DB analizzate in precedenza. Di seguito sono riportate le possibili modalità di vendita dei propri dati nel marketplace “databroker.global”.

- **Free:** permettere di offrire i propri dati gratuitamente. Dopo che l'utente è stato identificato, può accedere ai dati senza necessità di effettuare alcun pagamento. Questo modello è utilizzato principalmente da organizzazioni no-profit o per fornire un campione gratuito dei dati.
- **Prezzo fisso:** il venditore offre i dati a un prezzo fisso. In questo scenario, l'acquirente pagherà il prezzo richiesto, oppure contatterà il “fornitore” per ulteriori informazioni.
- **Prezzo Flessibile:** definisci un prezzo di vendita fisso, l'acquirente può acquistarlo immediatamente al prezzo appena definito o fare un'offerta privata a un prezzo inferiore. Quando viene ricevuta un'offerta, essa potrà essere rifiutata o accettata con un messaggio all'acquirente. Una volta accettata un'offerta, l'acquirente può finalizzare l'acquisto al prezzo concordato. Le offerte sono private, solo il venditore e l'acquirente sono a conoscenza del prezzo negoziato.
- **Prezzo di Mercato:** questo modello di vendita è suggerito per i venditori che non conoscono, ancora, il valore dei propri dati. Non viene definito il prezzo di vendita e l'acquirente può fare un'offerta. Come per il modello prezzo flessibile, il venditore sarà notificato e deciderà se accettare o rifiutare la proposta di prezzo dell'acquirente.
- **Prezzo in base alla domanda:** qui, non si definisce né un prezzo, né si permettono offerte, piuttosto, si aspetta prima una manifestazione di interesse da parte dell'acquirente interessato ai dati, e una volta che un acquirente di dati contatta il venditore, il prezzo può essere discusso così come i termini e le condizioni specifiche per quella particolare transazione. Questo modello di vendita è consigliato nel caso di mole di dati molto ampia con molte opzioni di personalizzazione a livello di contenuto dei dati o contrattuale.

3.9 CONCLUSIONI

L'elaborato ha presentato un'analisi sulle principali caratteristiche dei data broker, il cui settore è sommerso, ma tanto pervasivo quanto influente nella vita di tutti i giorni. L'attenzione delle autorità antitrust, e lentamente dei consumatori, sta iniziando a rivolgersi anche a tali figure.

Non sembrano esserci documenti di regolazione efficaci riguardanti il potere di mercato di queste figure.

La definizione del prezzo dei dati è sostanzialmente tenuta nascosta dai player, in quanto la maggioranza dei clienti pagano prezzi individuali su consulenze private.

L'elaborato ha evidenziato come i consumatori sono minacciati su due fronti: privacy ed estrazione del surplus. È stata messa in luce la questione dell'anonimizzazione. Essa viene comunemente presentata come la soluzione ai problemi di privacy; tuttavia, è facilmente aggirabile da parte dei data broker (e di conseguenza dai loro clienti); inoltre, la precisione delle informazioni a cui hanno accesso, sembrerebbe tendere ad una discriminazione del prezzo di primo ordine. Sono state, infine, citate due business nati di recente che si contrappongono al potere dei DB (Incogni e databroker.global)

Il settore è concentrato e la complementarità dei dataset fa presumere che ci si trovi nel caso di collusione tacita; tuttavia, l'analisi effettuata non è sufficiente per dimostrare ciò.

Per avvalorare la tesi appena descritta, si potrebbero confrontare i dataset di Equifax e Acxiom con quelli di Experian che, seguendo l'analisi effettuata per la costruzione di "Mosaic", presumibilmente, offre informazioni più dettagliate sulle tipologie e sui gruppi di nuclei familiari. Inoltre, Equifax offre informazioni sul credit scoring a granularità di singolo consumatore; tuttavia, con i dataset sample di AWS marketplace, non si trovano delle variabili che fungano da chiave per legare i dataset ai precedenti. A quel punto potranno essere ricoperte tutte le possibili combinazioni di granularità e rispondere all'esigenza di ogni cliente.

Infine, potrebbe essere utile pensare ad un modello di Hotelling lineare base, in cui le imprese sono i DB e i consumatori sono le imprese che acquistano i dati. La distanza, in questo caso, può essere intesa come la differenza di granularità desiderata e quella offerta dai DB, ovvero: se l'impresa i si trova esattamente dove è posizionato il DB_j , ciò significa che l'impresa i necessita di quella precisa granularità per vendere i propri prodotti; tuttavia, la maggioranza delle imprese ha bisogno di una granularità media che permetta di sfruttare la complementarità dei dataset; di conseguenza in tutte le posizioni diverse da quelle di DB_j ($\forall j$), le imprese devono comprare da tutti i data broker che massimizzeranno il profitto congiunto.

Il potere di mercato che si ottiene dai dati di cui si è proprietari offre un vantaggio in ogni settore e le aziende stanno cercando di ottenere sempre più dati. Un esempio potrebbe essere

Tesla che possiede le registrazioni video provenienti da tutte le sue vetture nel mondo. Si deduce che i dati così acquisiti possono essere riutilizzati per qualsiasi scopo e che, quindi, in un futuro non lontano Tesla possa entrare nel settore dei DB.

Bibliografia e sitografia

A. Muschalle, et al. 2012. "Pricing approaches for data markets," in *International Workshop on Business Intelligence for the Real-Time Enterprise*. 2012.

2024. About Experian. *Experian*. [Online] 2024. <https://www.experian.com/corporate/about-experian>.

Abrardi, Cambini, Carlo e Pino, Flavio. 2024. *Data Brokers Competition, Synergic Datasets, and Endogenous Information Value*. 2024.

ACCC. 2023. *Digital Platform Services Inquiry –March 2024 report on data brokers*. 2023.

ACXIAM. *Acxiom*. [Online] <https://www.acxiom.com/customer-data/>.

Avast. 2024. Data Brokers: Everything You Need to Know. *Avast*. [Online] 3 January 2024. <https://www.avast.com/c-data-brokers>.

AWS marketplace. 2024. Commercial Real Estate - AMP Insights (Demo/Sample). *AWS marketplace*. [Online] Equifax, 2024. https://aws.amazon.com/marketplace/pp/prodview-w62dypikxcrow?sr=0-1&ref_=beagle&applicationId=AWSMPContessa#dataSets.

—. **2024.** Consumer Credit Trends (Demo/Sample). *AWS marketplace*. [Online] Equifax, 2024. https://aws.amazon.com/marketplace/pp/prodview-tvppgxnhieqq?sr=0-1&ref_=beagle&applicationId=AWSMPContessa#offers.

—. **2024.** Equifax, Analytic Dataset. *AWS marketplace*. [Online] 2024. https://aws.amazon.com/marketplace/pp/prodview-vgmxxlm42lhmq?sr=0-11&ref_=beagle&applicationId=AWSMPContessa#offers.

—. **2024.** Germany - Statistical Data for Marketing Areas (refined Postcode) SAMPLE. *AWS*. [Online] 2024. https://aws.amazon.com/marketplace/pp/prodview-idkrmu5eg5dva?sr=0-12&ref_=beagle&applicationId=AWSMPContessa#offers.

—. **2024.** Germany - Statistical Data for Marketing Areas (refined Postcode) SAMPLE. *AWS marketplace*. [Online] Acxiom, 2024. https://aws.amazon.com/marketplace/pp/prodview-5f4o6htkujj4?sr=0-1&ref_=beagle&applicationId=AWSMPContessa.

—. **2024.** NYC Property Land Use Data (Pluto) - As of Sept 2019. *AWS marketplace*. [Online] Enigma, 2024. <https://us-east-1.console.aws.amazon.com/dataexchange/home?region=us-east-1#/subscriptions/prod-nlawf5p4gwnx2>.

—. **2024.** Real Property Insights (Demo/Sample). *AWS marketplace*. [Online] Equifax, 2024. https://aws.amazon.com/marketplace/pp/prodview-aixrwm3b4f5iy?sr=0-2&ref_=beagle&applicationId=AWSMPContessa#overview.

AWS marketplace, Acxiom consumer segmentation. 2024. Results. *Aws Marketplace*. [Online] Acxiom, 2024. https://aws.amazon.com/marketplace/search/results?prevFilters=%7B%22sr%22%3A%220-62%22%2C%22ref_%22%3A%22beagle%22%2C%22applicationId%22%3A%22AWSMPContessa%22%7D&searchTerms=acxiom+consumer+segmentation&pageSize=50.

AWS marketplace, WorldView Experian. 2024. Results. *AWS marketplace*. [Online] 2024. <https://aws.amazon.com/marketplace/search/results?prevFilters=%257B%2522id%2522%3A%2522dbf3ff8e-416c-41a0-9f44-1fbeb8c160a4%2522%257D&searchTerms=experian+world+view>.

Biden ha firmato la legge sul blocco di TikTok negli Stati Uniti: ora che succede? **Capone, Emanuele. 2024.** 2024, La Repubblica.

Burdova, Carly. 2020. Broker di dati: che cosa sono e come funzionano. *AVG*. [Online] 2020. <https://www.avg.com/it/signal/data-brokers>.

Casarella, Francesco. 2023. *Investing.com*. [Online] 2023. <https://it.investing.com/academy/statistics/google-azienda-fondatori-fatturato-e-curiosita/>.

Center for Strategic and International Studies (CSIS) . 2023. *Next Steps to Regulate U.S. Data Brokers and Their Interactions with Government Agencies*. 2023.

Center for Strategic and International Studies (CSIS). 2023. *Surveillance for Sale: The Underregulated Relationship between U.S. Data Brokers and Domestic and Foreign Government Agencies*. 2023.

Christl, Wolfie. 2017. *Corporate surveillance in everyday life*. 2017.

Commissione Europea. 2017. European Commission. *Guidelines on the right to "data portability" (wp242rev.01)*. [Online] 27 10 2017. <https://ec.europa.eu/newsroom/article29/items/611233>.

—. **2020.** Summary report of the public consultation on the European strategy for data. *European Commission*. [Online] 2020. <https://digital-strategy.ec.europa.eu/en/library/summary-report-public-consultation-european-strategy-data>.

CORELOGIC. 2024. CoreLogic. *CoreLogic*. [Online] 2024. <https://www.corelogic.com/>.

—. **2024.** CoreLogic Knows Property. *CoreLogic*. [Online] 2024. <https://www.corelogic.com/>.

Crain, Matthew. 2017. *The limits of transparency: Data brokers and commodification*. s.l. : CUNY Queens College, 2017.

Databroker.global. 2024. Documentation. *Databroker.global*. [Online] 2024. <https://www.databroker.global/documentation>.

David Bounie, Dubus , Antoine e Waelbroeck, Patrick. 2021. *Selling strategic information in digital competitive markets*. 2021.

2024. Empowering Intelligent Decision Making. *Illion*. [Online] 2024. <https://www.illion.com.au/who-we-are/>.

Evan Selinger, Jules Polonetsky, Omer Tene. 2018. *The Cambridge Handbook of Consumer Privacy*. s.l. : Cambridge University Press, 2018.

EXPERIAN. 2024. Credit 3D. *Experian*. [Online] 2024. <https://www.experian.it/business/analytics-and-decisioning/dati-alternativi/credit-3d>.

Experian. 2019. *Mosaic USA E-Handbook*. 2019.

FBI. 2020. Chinese Military Hackers Charged in Equifax Breach. *FBI*. [Online] 2020. <https://www.fbi.gov/news/stories/chinese-hackers-charged-in-equifax-breach-021020>.

Federal Commision Trade (FTC). 2014. *Data Brokers, a Call for Transparency and Accountability*. 2014.

FTC. 2024. Equifax Data Breach Settlement: What You Should Know. *Federal trade commission*. [Online] 2024. <https://consumer.ftc.gov/consumer-alerts/2019/07/equifax-data-breach-settlement-what-you-should-know>.

Gu, Yiquan, Madio, Leonardo e Reggiani, Carlo. 2021. *Data Brokers Co-Opetition*. 2021.

Harmon, R., et al. 2009. *Pricing strategies for information technology services: A value-based approach*. 2009.

ILLION. 2024. Illion, Empowering Intelligent Decision Making. *Illion*. [Online] 2024. (Empowering Intelligent Decision Making, 2024).

Incogni. 2024. *Incogni*. [Online] 2024. <https://incogni.com/>.

—. **2024.** About us- Incogni. *Incogni*. [Online] 2024. <https://incogni.com/about-us>.

Krämer, Schnurr, Daniel e Haberer, Bastian . 2018. *Do Consumers Benefit from Selling Their Data? The Economic Impact of Personal Data Brokers on Digital Markets*. 2018.

Langfelder, Natasia. 2023. data axle. *The complete guide to third-party data*. [Online] 2023. <https://www.data-axle.com/resources/blog/the-complete-guide-to-third-party-data/>.

Liang, Fan , et al. 2018. *A Survey on Big Data Market: Pricing, Trading and Protection*. 2018.

Maia Research. 2023. Global Data Broker Service Market Research Report 2023-Competitive Analysis, Status and Outlook by Type, Downstream Industry, and Geography, Forecast to 2029. *Market Research*. [Online] 2023. <https://www.marketresearch.com/Maia-Research-v4212/Global-Data-Broker-Service-Research-33786522/>.

Mcdonald, Cranor e Faith Cranor, Lorrie. 2008. *The cost of reading privacy policies*. 2008.

Moody's. 2024. *Orbis*. [Online] 2024. https://www.moody's.com/web/en/us/capabilities/company-reference-data/orbis.html?cid=ppc-gglds-16892&gad_source=1&gclid=CjwKCAjwgdayBhBQEiwAXhMxtjvpqmuTi4qs0n264ytOmJ1HTNkeN2OI0y7jbS19f7XvyXSA76S7yBoCDsgQAVD_BwE&gclsrc=aw.ds.

1990. *Moore v. Regents of University of California*. 51 Cal.3d 120, s.l. : Supreme Court of California, 9 July 1990.

ORACLE. 2024. *Oracle*. [Online] 2024. <https://www.oracle.com/it/corporate/>.

Oracle, fatturato record di 50 miliardi di dollari nell'anno fiscale 2023. **La Stampa. 2023.** 2023, La Stampa.

PitchBook. 2024. DataCoup Overview. *PitchBook*. [Online] 2024. <https://pitchbook.com/profiles/company/95937-58#faqs>.

Reviglio, Urbano. 2022. *The untamed and discreet role of data brokers in surveillance capitalism: a transnational and interdisciplinary overview*. 2022.

Rieke, Aaron. 2016. *Data Brokers in an open society*. 2016.

Ruscheimer, Hannah. 2023. *Data Brokers and European Digital Legislation*. 2023.

Sturabotti, Alessandra. 2021. Commercio e divieto di vendita dei dati personali: i due opposti approcci. *Agenda digitale*. [Online] 27 October 2021. <https://www.agendadigitale.eu/sicurezza/privacy/commercio-e-divieto-di-vendita-dei-dati-personali-approcci-speculari-apologeti-contro-proibizionisti/>.

Swendeman, Dave. 2023. First-Party Data | Definition, Examples, Segmentation. *daasity*. [Online] 2023. <https://www.daasity.com/post/first-party-data>.

Venkatadri, Giridhari. 2019. *Investigating sources of PII used in Facebook's targeted advertising*. 2019.

2023. What are data brokers? *INCOGNI*. [Online] 2023. <https://incogni.com/data-brokers>.