

POLITECNICO DI TORINO

Corso di Laurea Magistrale in Ingegneria Gestionale

Anno Accademico 2019/2020



Big Data, Concorrenza, Privacy e loro interdipendenza

Relatore:

Prof. Carlo Cambini

Candidato:

Annunziata Donato

Dicembre 2020

Indice

Introduzione	1
Big Data nell'era digitale.....	3
1.1 Cosa si intende per Big Data	4
1.2 Caratteristiche dei Big Data: le 5 V.....	6
1.2.1 Volume	6
1.2.2 Varietà	7
1.2.3 Velocità	8
1.2.4 Veridicità	8
1.2.5 Valore	9
1.3 La Value Chain dei Big Data	9
1.3.1 Raccolta	10
1.3.2 Archiviazione	13
1.3.3 Analisi	14
1.3.4 Utilizzo	14
1.4 L'ecosistema dei Big Data.....	15
1.5 Categorie di dati	18
1.5.1 Classificazione in base all'origine dei dati.....	18
1.5.2 Classificazione in base al grado di identificazione personale	22
1.5.3 Classificazione in base alla struttura dei dati	24
Big Data: fonte di potere di mercato?	29
2.1 Stabilire un "Data Power"	31
2.1.1 Barriere all'ingresso ai mercati dei Big Data	32
2.2 Uso dei Big Data e Vantaggio Competitivo	34
2.2.1 Servizi personalizzati	34
2.2.2 Raccomandazioni personalizzate.....	36
2.2.3 Targeted advertising	37
2.3 Fattori che facilitano il potere di mercato.....	40
2.3.1 Economie di scala e di scopo	40
2.3.2 Feedback loops	41
2.3.3 Esternalità di rete.....	42
2.4 Fattori che attenuano il potere di mercato: il multi-homing.....	44
Big Data e Piattaforme Online.....	46
3.1 Definizione e tipi di piattaforme multi-versante	46

3.2	Caratteristiche delle piattaforme multi-versante	48
3.2.1	Effetti di rete.....	48
3.2.2	Pricing asimmetrico.....	49
3.2.2.1	Modello di Parker e Van Alstyne (2005)	52
3.2.2.2	Modello di Rochet e Tirole (2006).....	54
3.3	Impatto economico delle piattaforme online.....	55
3.4	Le piattaforme di attenzione.....	57
3.4.1	Ruolo dei Big Data: il costo del “gratuito”	58
3.4.2	Oligopolio dell’attenzione.....	59
3.4.3	Il modello di Prat e Valletti (2018)	61
3.4.2.1	Ipotesi del modello	62
3.4.2.2	Analisi del modello	62
3.4.2.3	Conclusioni del modello.....	68
	Big Data: tra concorrenza e privacy	69
4.1	Privacy e sua tutela nei mercati digitali.....	70
4.1.1	Atteggiamiento dei consumatori: il “paradosso della privacy”	70
4.1.2	La privacy nell’economia “zero-price”	71
4.1.3	Regolare la privacy online: il GDPR.....	73
4.1.3.1	I principi fondamentali del GDPR.....	74
4.1.3.2	Il ruolo dell’interessato al trattamento.....	75
4.1.3.3	Il diritto alla portabilità dei dati.....	75
4.1.3.4	Responsabilità congiunta del titolare e del responsabile del trattamento... ..	76
4.2	Impatto del GDPR sulla concorrenza	76
4.2.1	Effetti indesiderati del GDPR: Studio empirico Batikas et al. (2020)	77
4.2.1.1	Contesto dell’analisi	77
4.2.1.2	Dati dell’analisi	79
4.2.1.3	Modello econometrico di base.....	80
4.2.1.4	Richieste a domini di terzi-parti	80
4.2.1.5	Struttura del mercato e concorrenza	82
4.2.1.6	Conclusioni dello studio	88
4.3	Il caso: Facebook nel mirino del Bundeskartellamt	89
4.3.1	La privacy è una questione di concorrenza?.....	89
4.3.2	L’indagine del Bundeskartellamt	91
4.3.2.1	Analisi del contesto	92
4.3.2.2	La raccolta dati off-platform sotto la lente del Bundeskartellamt.....	94
4.3.2.3	Definizione del mercato rilevante	95

4.3.2.4 Analisi del potere di mercato.....	96
4.3.3 La condotta di Facebook è in violazione della legge?.....	97
4.3.3.1 Esame della compatibilità delle pratiche di Facebook con il GDPR.....	97
4.3.3.2 Esame della condotta sotto il profilo della concorrenza.....	99
4.3.4 Conclusione del caso e decisione dell'antitrust tedesca.....	100
Conclusioni	102
Bibliografia	104

Elenco delle figure

Figura 1.1 Trend delle query avente ad oggetto il termine 'Big Data' nel motore di ricerca di Google (originato con Google Trend).....	4
Figura 1.2 Dimensione annua della Datasfera Globale [7].....	7
Figura 1.3 Stima dei dati creati su internet in un minuto. Fonte: www.statista.com	8
Figura 1.4 Le principali fasi della catena del valore dei Big Data.....	9
Figura 1.5 Tecnologie di tracciamento sul web [11].....	10
Figura 1.6 Ecosistema dei Big Data [18]	16
Figura 1.7 Tipi di dati in base al modo in cui sono stati generati.....	19
Figura 1.8 Tassonomia dei dati raccolti e creati da un fornitore di servizi online.....	20
Figura 1.9 L'impatto dei modelli di ricavo sull'uso dei dati [12].....	21
Figura 1.10 Tipi di dati dei consumatori personali e non personali.....	22
Figura 1.11 Principali differenze tra i dati strutturati e non strutturati.....	24
Figura 2.1 Quote di mercato globali per società (2019). Fonte: GSStatCounter, eMarkete, Statista.....	29
Figura 2.2 Forme diverse di targeted advertising [23].	38
Figura 2.3 Dal Data Power al Vantaggio Competitivo.....	39
Figura 2.4 Il ciclo di feedback degli utenti [28]	41
Figura 2.5 Il ciclo di feedback di monetizzazione [28].	42
Figura 3.1 Effetti di rete diretti e indiretti in un mercato a due lati.....	48
Figura 3.2 Curve di domande e profitti con pricing tradizionale.	51
Figura 3.3 Curve di domande e profitti con pricing asimmetrico.....	51
Figura 3.4 Curve di reazione in presenza di livelli diversi di esternalità indiretta.....	53
Figura 3.5 Società di piattaforme classificate in base al tipo [48]	57
Figura 3.6 Scambi di valore nella fornitura di servizi a costo monetario zero [51]	59

Figura 3.7 Quote di mercato di Google e Facebook rispetto ai ricavi globali della pubblicità online e della pubblicità sui media in generale. Fonte: www.statista.com	60
Figura 3.8 Alberi di gioco per la concorrenza tra un'azienda incumbent e un'azienda entrant per l'acquisto di un annuncio pubblicitario [49]	64
Figura 4.1 Esempio di richieste di terze parti [64]	77
Figura 4.2 Andamento numero medio di tracker di terze parti sui siti web con pubblico UE e US. [65].....	78
Figura 4.3 Andamento site reach, ossia della % dei siti web in cui il tracker è presente [65].....	79
Figura 4.4 Log del numero di richieste a domini di terze parti [64].	80
Figure 4.5 Variazione del numero di domini di terze parti richiesti, previsione out-of-sample [64].....	81
Figura 4.6 Cambiamento nella struttura di mercato del settore delle tecnologie di tracking web [64]	82
Figura 4.7 Winners e losers 2 settimane prima/dopo l'entrata in vigore del GDPR [64]....	86
Figura 4.8 Winners e losers 2 settimane prima/dopo l'entrata in vigore del GDPR [64].....	86
Figura 4.9 Sovrapposizione tra protezione dei dati, concorrenza e protezione dei consumatori [51].	89
Figura 4.10 Mappa mondiale dei social networks. Fonte: vincos.it	92
Figura 4.11 Online tracking by Google e Facebook [51]	93
Figura 4.12 Lati della piattaforma multi-versante di Facebook. [70]	95

Elenco delle tabelle

Tabella 2.1 Barriere all'ingresso nella supply chain dei dati [14].....	32
Tabella 3.1 Esempi di mercati two-sided [46].....	50
Tabella 3.1 Condizioni di equilibrio delle aste nei tre casi [49].....	67
Tabella 4.1 Cambiamento nel numero di richieste a domini di terze parti.....	81
Tabella 4.2 Cambiamenti nella struttura di mercato del settore della tecnologia web.....	83
Tabella 4.3 Cambiamento nelle quote di mercato dei fornitori di tecnologia web.....	84
Tabella 4.4 Cambiamenti nelle quote di mercato, per sottomercato, Google vs tutte le altre aziende.....	87
Tabella 4.5 Cambiamenti nelle quote di mercato, per sottomercato, Google e Facebook contro tutte le altre aziende.....	88

Introduzione

Il presente lavoro di tesi affronta un tema di grande attualità che è stato, ed è destinato ancora ad essere, al centro degli interessi non solo degli studiosi e delle autorità competenti, ma di tutta l'opinione pubblica: i Big Data e le implicazioni che la loro raccolta e il loro uso hanno in termini di concorrenza e di privacy.

È infatti ormai appurato che, a partire soprattutto dall'ultimo decennio, i cambiamenti imposti dall'innovazione tecnologica hanno generato un livello senza precedenti di raccolta ed elaborazione di dati, destinato a subire un'ulteriore espansione per via delle nuove applicazioni dell'Internet of Things, della robotica, della realtà aumentata.

In questo contesto sono in particolare i modelli di business delle piattaforme online, le quali si basano su una vasta raccolta ed elaborazione dei dati prodotti *dalle e sulle persone*, nonché sulla loro elaborazione in tempo quasi reale, a sollevare una serie di interrogativi ineludibili. Sebbene infatti essi abbiano da un lato permesso di offrire servizi innovativi, personalizzati e spesso a prezzo zero, e abbiano dunque apportato notevoli vantaggi - quanto meno se si considera il benessere a breve termine - per i consumatori, dall'altro il fatto che possano disporre di particolari condizioni attraverso cui il possesso e l'utilizzo di Big Data possa conferire loro un potere di mercato difficile da scardinare è fonte di non poche preoccupazioni. Ad alimentarle, inoltre, vi è l'evidenza del successo commerciale dirompente di una manciata di aziende che costituiscono oramai dei gateway fondamentali ai servizi che il mondo Internet mette a disposizione.

La loro capacità di accedere ad una quantità di dati incomparabile per i concorrenti consente loro di ricavare da essi delle previsioni molto più sofisticate e accurate delle preferenze dell'utente e di monetizzare in modo più efficace attraverso l'offerta mirata di servizi presso gli utenti che più li apprezzano.

Seppur, ovviamente, non tutti i Big Data si riferiscono a dati personali, per molti servizi online che vengono presentati o percepiti come "gratuiti", e che saranno oggetto di questa discussione, le informazioni personali funzionano come una sorta di valuta indispensabile per potervi accedere. Ecco perché l'attenzione alla questione dei Big Data non può riguardare soltanto la sua implicazione in termini di concorrenza, ma deve preoccupare anche il potenziale discriminatorio che dal loro utilizzo, anche rispetto a dati non identificativi, aggregati, può nascere per effetto di profilazioni sempre più puntuali e analitiche e dunque la sua implicazione pure in termini di privacy. Lo sfruttamento dei dati personali può anche risultare da una coercizione economica, sulla base della dipendenza dalle risorse o del lock-in dell'utente, che non ha altra scelta, al fine di godere del consumo di un servizio specifico fornito dal titolare del trattamento o dal suo ecosistema, se non quella di acconsentire alla raccolta e all'utilizzo dei suoi dati.

Pertanto, questa tesi affronta anche il tema dei modi in cui il diritto della concorrenza potrebbe essere coinvolto in comportamenti di sfruttamento che portano a danni alla privacy. A tal proposito, il caso Bundeskartellamt versus Facebook relativo all'eccessiva raccolta dati da parte di quest'ultimo rappresenta uno dei primissimi tentativi in cui si dimostra come il trattamento dei dati personali possa essere non solo, come

tradizionalmente è stato, il campo di applicazione di leggi e regolamenti volti a proteggere la privacy dei cittadini ma debba essere integrato con il diritto della concorrenza.

In particolare, nel *primo capitolo*, si fornisce un'introduzione all'argomento dei Big Data, considerando le sue principali caratteristiche, la catena del valore relativa nonché le diverse categorie di dati, alcune delle quali di particolare rilievo relativamente al tema della tutela della privacy.

Nel *secondo capitolo* si discute del ruolo dei Big Data come possibile base di un potere di mercato duraturo, evidenziando i modi attraverso i quali le aziende dotate di un potere dei dati ("Data Power") possono trasformarlo in un vantaggio competitivo e quali fattori possono poi promuoverne la sua persistenza.

Nel *terzo capitolo*, l'attenzione è rivolta alle piattaforme online, aziende che fanno della raccolta e dell'uso dei dati degli utenti uno dei punti chiave della loro strategia. Si considereranno in particolare le piattaforme di attenzione, la loro condizione oligopolistica e gli effetti di questa sulla concorrenza e sul benessere del consumatore.

Infine, nel *quarto e ultimo capitolo*, si affronterà il tema dell'interdipendenza tra il diritto della concorrenza e la tutela della privacy prendendo in considerazione il caso dell'autorità garante della concorrenza tedesca contro Facebook, nonché l'impatto che il GDPR, l'attuale regolamento generale per la protezione dei dati europeo, può determinare in termini di concentrazione delle imprese operanti nel mercato dei dati.

CAPITOLO 1

Big Data nell'era digitale

Dall'inizio degli anni 2000, un nuovo fenomeno in rapida crescita sta cambiando le nostre società: la digitalizzazione. L'adozione in massa di servizi digitali connessi da parte di consumatori, imprese e istituzioni è, infatti, emersa come un motore economico chiave che ha trasformato le nostre vite con una scala e una velocità inedita, offrendo immense opportunità e al contempo sfide spaventose. I servizi online stanno diventando sempre più centrali nel modo in cui le persone soddisfano le loro esigenze quotidiane di comunicazione, di acquisto, di informazione, di intrattenimento e sono sempre di più sostituiti di canali tradizionali. È evidente pure una tendenza generale di una connessione informatica diffusa di persone, cose, luoghi e processi¹ che contribuisce ad aumentare in modo esponenziale la disponibilità di dati digitali.

I dati, e soprattutto la conoscenza che è possibile estrarne dalla loro analisi, hanno sempre avuto un gran valore nelle attività economiche e sociali, ma ciò che c'è di nuovo rispetto al passato è che mentre prima del mondo si offrivano rappresentazioni analogiche che potevano poi essere trasformate in sequenze di uno e zero, oggi è aumentata in modo significativo la capacità delle istituzioni, imprese e individui, di “*datificare*” la realtà, ossia di disporre di rappresentazioni degli eventi e dei comportamenti umani che nascono già digitali. In particolare, negli ultimi anni, più che parlare di semplici o singoli dati, si è soliti riferirsi ad un concetto molto più complesso e ampio riconducibile al termine di **Big Data**, proprio in virtù dell'ingente quantità di dati disponibili all'interno del nuovo ecosistema digitale, prodotti ad alta velocità e provenienti da una moltitudine di fonti, la cui gestione e analisi richiedono nuovi, più performanti e intelligenti strumenti in termini di processori e algoritmi.

Queste enormi quantità di dati sono utilizzate in modo estensivo per contribuire a migliorare i servizi e guidare l'innovazione, servire la pubblicità mirata e sviluppare tecniche di intelligenza artificiale (AI) e di machine learning (ML). Svariati sono i settori che sfruttano tale risorsa in diverse applicazioni: così come le aziende raccolgono e utilizzano questi dati per supportare modelli di business personalizzati per la fornitura di prodotti e servizi, i governi utilizzano i dati personali per fornire servizi pubblici essenziali in modo più efficiente ed efficace, mentre ancora ricercatori accelerano ad esempio lo sviluppo di nuovi farmaci e protocolli di trattamento.

¹ Il traffico Internet totale ha registrato una crescita vertiginosa negli ultimi due decenni. Lo [studio Cisco Visual Networking Index™ 2018](#) sostiene che più di 20 anni fa, nel 1992, le reti Internet globali trasportavano circa 100 GB di traffico al giorno. Dieci anni dopo, nel 2002, il traffico Internet globale ammontava a 100 gigabyte al secondo (GBps). Nel 2017, il traffico Internet globale ha raggiunto più di 45.000 GBps e si prevede che nel 2022 supererà i 150.000 GBps, alimentato da sempre più persone che arrivano online per la prima volta e dall'espansione dell'IoT.

È dunque facile comprendere che i dati sono uno dei principali motori della creazione di valore nell'era digitale e, inoltre, l'aumento vertiginoso della loro importanza si deve fondamentalmente a due tendenze tecnologiche relativamente recenti: il progresso tecnologico e lo sviluppo di sofisticate tecniche analitiche. Il progresso tecnologico ha, infatti, ridotto drasticamente i costi della raccolta, dell'archiviazione e dell'uso di dati che, a causa delle attività economiche e sociali sempre più digitalizzate, vengono costantemente prodotti. Lo sviluppo di sofisticate tecniche analitiche ha, invece, permesso di ottenere gradi avanzati di elaborazione dei dati grazie ai quali è possibile estrarre maggior valore da essi. L'importanza di questi due fattori può essere illustrata dal fatto che i dati sono al centro dei modelli di business altamente redditizi delle più grandi aziende del mondo, tanto è vero che non è infrequente l'espressione *"i dati sono il nuovo petrolio"*², sebbene differentemente dal petrolio, risorsa finita e non riutilizzabile, i dati possono essere - tenuti ovviamente in considerazione la proprietà e i diritti di accesso- infiniti e riutilizzabili.

1.1 Cosa si intende per Big Data

Il termine Big Data è ormai da anni una parola di tendenza sia nel mondo accademico che in quello industriale, tanto è vero che, se si considera la frequenza delle query di ricerca sul Web (Figura. 1.1) come un'approssimazione dell'attenzione intorno a questo fenomeno, si può osservare come la sua popolarità sia cresciuta enormemente a partire dal 2011.

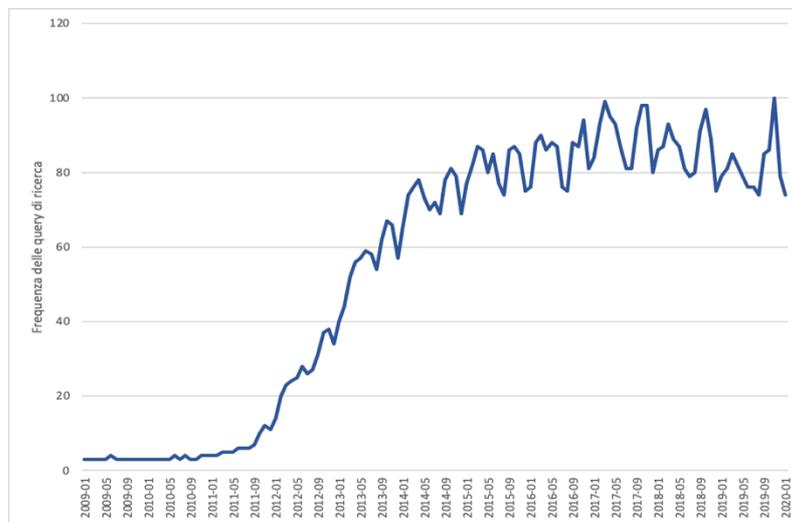


Figura 2.1: Trend delle query avente ad oggetto il termine 'Big Data' nel motore di ricerca di Google (originato con Google Trend)

Tuttavia, il grado di popolarità di questo fenomeno non è stato accompagnato da uno sviluppo razionale di un vocabolario accettato. Infatti, lo stesso termine "Big Data" è stato usato con diverse e incoerenti accezioni e manca di una definizione formale e di omogeneità. Le definizioni esistenti forniscono prospettive molto diverse: Big Data è considerato, di

² Si veda <https://www.economist.com/leaders/2017/05/06/the-worlds-most-valuable-resource-is-no-longer-oil-but-data>

volta in volta, un termine che descrive un fenomeno sociale, il patrimonio informativo, i set di dati, le tecnologie di archiviazione, le tecniche di analisi, i processi e le infrastrutture.

Nonostante però l'uso del termine "Big Data" sia spesso vago e manchi di precisione, le definizioni più frequentemente utilizzate si riferiscono di solito (1) alla grande dimensione del set di dati e (2) alla necessità di utilizzare potenza di calcolo su larga scala e software e metodi non standard per estrarre valore dai dati in un ragionevole lasso di tempo.

Una semplice definizione può ricondursi a Manyika et al (2011) secondo cui "*Con 'Big Data' ci si riferisce a set di dati la cui dimensione va oltre la capacità dei tipici strumenti software di database di catturare, archiviare, gestire e analizzare*" [1]. Si tratta volutamente di una definizione in cui il concetto stesso di Big Data è considerato come un obiettivo in movimento e non da vincolare a delle soglie convenzionali che un dataset deve rispettare per essere considerato 'Big'. Questo perché il volume dei dati catturato secondo la definizione odierna può cambiare nel tempo con l'aumentare delle capacità di archiviazione e di analisi; ciò che oggi, infatti, è possibile con i petabyte, solo pochi anni fa non lo era con i terabyte.

Per definire ulteriormente la dimensione del fenomeno, è utile fare riferimento al report IDC "*The Digital Universe in 2020*", che afferma che i Big Data rappresentano: "*una nuova generazione di tecnologie e architetture, progettate per estrarre economicamente valore da volumi molto grandi di una grande varietà di dati, consentendo la cattura, la scoperta e/o l'analisi ad alta velocità*" [2].

Infine, alla ricerca di una definizione più completa del fenomeno, De Mauro et al (2016), discutendo le proposte esistenti e indagandone le caratteristiche principali, hanno cercato di sintetizzare un'espressione che rappresentasse formalmente l'essenza del fenomeno:

"I Big Data sono il patrimonio informativo caratterizzato da un volume, una velocità e una varietà di dati così elevati da richiedere una tecnologia e metodi analitici specifici per la loro trasformazione in valore".

Tale analisi congiunta delle definizioni esistenti per i Big Data e dei principali temi di ricerca in letteratura ha permesso, dunque, di concludere che il nucleo del concetto di Big Data comprende i seguenti aspetti [3]:

- *Volume, Velocità e Varietà*, per descrivere le caratteristiche dell'informazione;
- *Tecnologia e Metodi di analisi*, per descrivere i requisiti necessari per fare un uso corretto di tali informazioni;
- *Valore*, per descrivere la trasformazione delle informazioni in intuizioni che possono creare valore economico per le aziende e la società.

1.2 Caratteristiche dei Big Data: le 5 V

Un utile modo di caratterizzare gli attributi chiave dei Big Data solitamente è quello di fare riferimento alle “5 V”: il *Volume*, la *Velocità*, la *Varietà*, la *Veridicità* e il *Valore*. Si tratta di caratteristiche la cui convergenza aiuta sia a definire sia a distinguere i Big Data dai dati in generale.

Tale caratterizzazione si ispira alla definizione delle "3 V" introdotta originariamente da Laney (2001), che considerava come principali dimensioni dei Big Data il *volume* dei dati, la *velocità* con cui i dati sono raccolti, utilizzati e diffusi e la *varietà* di informazioni aggregate [4].

Successivamente, la V supplementare di *veridicità* è stata inclusa da IBM per sottolineare l'importanza di indirizzare e gestire l'incertezza insita in alcuni tipi di dati [5]. Infine, a completare ulteriormente il quadro si può includere il *valore*, sia come dimensione base che come conseguenza delle altre quattro caratteristiche.

1.2.1 Volume

Il *Volume* è forse la caratteristica più associata ai Big Data e si riferisce alle grandi quantità di dati che le organizzazioni stanno cercando di sfruttare per migliorare il loro processo decisionale.

Negli anni sono stati principalmente due i fattori chiave che hanno contribuito all'aumento esponenziale del volume di dati raccolti e sfruttati [6]:

- La riduzione dei costi per la raccolta, la memorizzazione, l'elaborazione e l'analisi dei dati.
- La crescente attività online dei consumatori, spinta da un maggiore accesso a Internet ad alta velocità, nonché da un maggior numero di beni e servizi online e connessi, compresi quelli forniti dai dispositivi dell'IoT.

Si stima che il volume dei dati disponibili, come conseguenza dell'ubiquità dell'attività online e di Internet, continuerà a crescere ulteriormente. IDC (International Data Corporation) prevede infatti che entro il 2025 la *Datasfera Globale*³ arriverà a 163 zettabyte⁴, con una crescita di volume di circa dieci volte rispetto a quello registrato nel 2016 (Figura 1.2). Saranno, in particolare, gli endpoint, ovvero tutti i dispositivi ai margini

³ Per *Datasfera Globale* si intende l'insieme dei dati prodotti a livello mondiale e prevede tre categorie di località per essi: il *Core* che comprende i grandi data center inclusi il cloud pubblico e privato, l'*Edge* ovvero l'insieme di computer e apparecchiature, al di fuori del nucleo, in grado di funzionare all'interno dell'impresa, come ad esempio sale server, server sul campo e data center più piccoli situati a livello regionale per tempi di risposta più rapidi, ed infine gli *Endpoint* ossia tutti i dispositivi ai margini della rete, compresi PC, telefoni, macchine fotografiche, auto collegate, wearables e sensori.

⁴ Un zettabyte equivale a 1 trilardo (1021) byte. Per avere un'idea pratica della quantità di dati che 1 zettabyte esprime basti pensare che questo corrisponde a circa 180 milioni di volte le documentazioni conservate nella Biblioteca del Congresso di Washington, <https://it.wikipedia.org/wiki/Zettabyte>

della rete, compresi i PC, i telefoni, le macchine fotografiche, le auto connesse, i wearables e i sensori, a dare in percentuale maggiore un notevole impulso alla creazione totale dei dati [7].

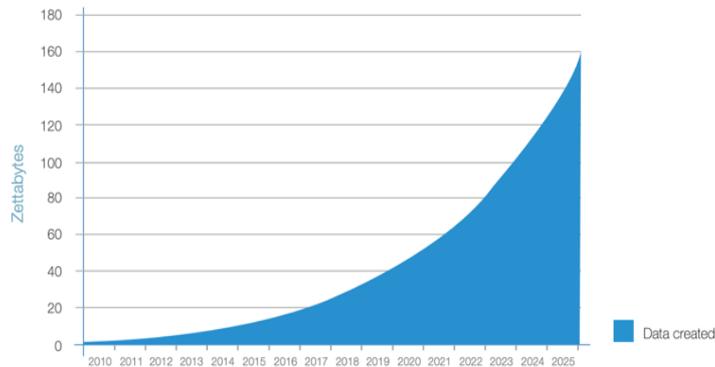


Figura 1.2: Dimensione annua della Datasfera Globale [7]

1.2.2 Varietà

La seconda caratteristica associabile ai Big Data è la *varietà*, da intendersi in riferimento sia ai diversi tipi che alle diverse fonti di dati. L'esplosione dell'uso di dispositivi intelligenti, dei social network, dei sensori contribuisce a costituire un complesso insieme di fonti, interne ed esterne all'impresa, che genera dati in innumerevoli forme, tra cui dati strutturati, semi-strutturati e non strutturati⁵.

L'odierna disponibilità preponderante di questi ultimi contribuisce ulteriormente ad incrementare la complessità della gestione dell'eterogeneità. Tale tipologia - che include dati quali immagini, foto, testi, audio, video - non è infatti organizzata in strutture composte da righe e colonne che possono essere facilmente ordinate ed elaborate secondo tecniche riferite a database relazionali (RDBMS, fogli di calcolo, data warehouse, Customer Relationship Management System, ecc.), dalle quali è sicuramente più facile estrarre valore grazie all'utilizzo di tecniche consolidate. Al contrario, trattandosi di dati che non sono organizzati secondo una struttura definita, necessitano di tecniche molto sofisticate capaci di estrarre informazioni da essi.

Oggi più che mai, sia per la presenza di diverse fonti di dati, sia per le grandi opportunità derivanti dalla possibilità di combinare dati di formati diversi, la varietà è una delle caratteristiche più distintive di molti modelli di business data-driven, quale quello delle piattaforme online. Lo confermano i risultati di una ricerca condotta da NVP (New Vantage Partners) da cui emerge che il 40% delle aziende intervistate - appartenenti alla classifica Fortune 1000 - fa riferimento alla varietà come il principale motore nelle scelte di investimento per la gestione dei Big Data, rispetto al 14,5% che cita il volume e al 3,5% che indica la velocità [8] [9].

⁵ Si consideri a tal proposito la sezione 1.5.3.

1.2.3 Velocità

La terza caratteristica, ossia la *velocità*, fa riferimento sia alla velocità di generazione e accesso ai dati, sia alla velocità di elaborazione ed analisi dei dati. Anche questa ulteriore dimensione dei Big Data, così come le precedenti, ha subito una crescita tale da rendere le tecniche tradizionali di raccolta ed elaborazione dei dati non più efficaci.

La Figura 1.3, oltre a fornire una sintesi di quanti dati vengono prodotti quindi acquisiti su internet in soli 60 secondi, dà un'idea del funzionamento congiunto delle tre principali caratteristiche dei Big Data (volume, varietà e velocità).

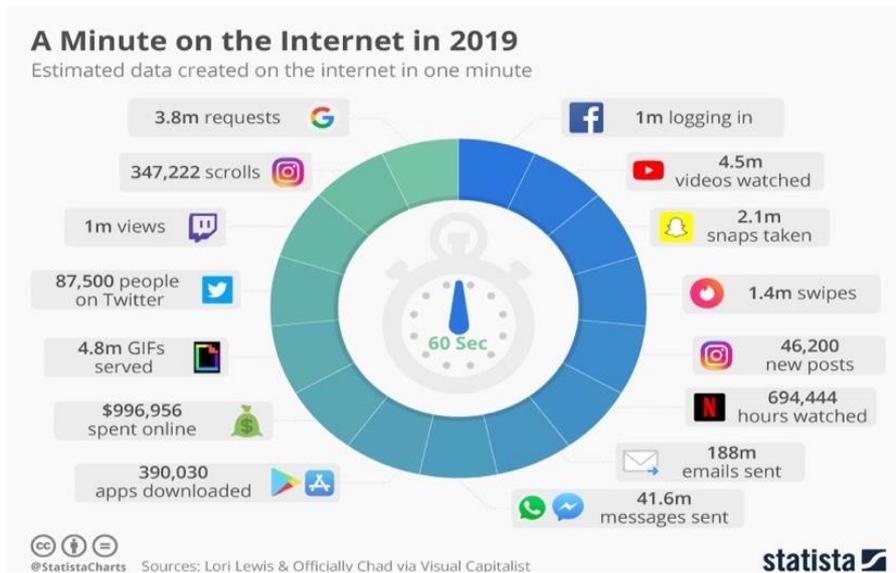


Figura 1.3: Stima dei dati creati su internet in un minuto. Fonte: www.statista.com

A contribuire ad una velocità più elevata è anche la necessità per processi sensibili al tempo, come il marketing o l'advertising "istantaneo", di analizzare i dati in tempo reale e di prendere decisioni ad un ritmo sempre crescente, in modo tale che questi siano di valore per l'impresa e riescano a garantire le migliori performance.

È infine importante sottolineare che se è vero che in alcuni contesti il valore dei dati non è deperibile col tempo, è altrettanto vero che per molte opportunità di business legate alla capacità di sfruttare rapidamente e tempestivamente i dati disponibili, l'analisi in tempo reale è ormai un requisito fondamentale. Tuttavia, questo aspetto richiede competenze, infrastrutture tecnologiche e soluzioni software sofisticate.

1.2.4 Veridicità

Un'ulteriore "V" idonea a caratterizzare i Big Data è sicuramente la *veridicità*, ovvero la qualità e significatività dei dati raccolti o elaborati, soprattutto alla luce dell'attuale aumento esponenziale delle tre dimensioni precedentemente definite e in virtù della sempre maggiore diffusione dei processi di machine learning.

La qualità e l'integrità dei dati costituisce un requisito imprescindibile- tanto è che tra gli esperti del settore si usa dire "*Bad data is worse than no data*" - per far sì che le

organizzazioni, lavorando con dati “veritieri” diano vita ad analisi utili e al contempo affidabili. Per raggiungere tale obiettivo spesso si rendono necessarie tecniche di preprocessing dei dati raccolti volte ad eliminare i dati non significativi e/o non integri.

1.2.5 Valore

Le quattro caratteristiche precedenti servono come base per la “V” finale: il *valore* che i dati assumono quando sono elaborati e analizzati, soprattutto in riferimento all'accuratezza delle previsioni e alla capacità di generare nuove opportunità di business impensabili in precedenza.

La capacità di estrarre valore dai Big Data si deve principalmente al proliferare di tecniche analitiche sempre più sofisticate che consentono, tramite l'estrazione di informazioni dai dati, di migliorare il processo decisionale e le prestazioni, contribuendo all'efficienza e alla qualità dei processi e qualificando l'offerta di beni e/o servizi, in particolare, in termini di innovazione e personalizzazione [10].

È interessante notare, poi, che soprattutto nel panorama digitale il valore che i Big Data possono generare è talmente elevato che le aziende a volte, in cambio della raccolta e dell'utilizzo dei dati dei consumatori, sono disposte a fornire loro servizi “gratuiti”.

In definitiva, i Big Data sono una combinazione di queste caratteristiche che confermano una dinamica evolutiva molto rapida e complessa del fenomeno, ma che potrebbero creare un'opportunità di guadagno per le organizzazioni e un vantaggio competitivo nell'odierno mercato digitalizzato.

1.3 La Value Chain dei Big Data

Nel processo di estrazione di valore dai Big Data è possibile individuare una sequenza di attività principali: la *Raccolta*, l'*Archiviazione*, l'*Analisi* e infine l'*Utilizzo* (Figura 1.4). Dalla raccolta all'uso, infatti, l'insieme di dati procede attraverso diverse fasi interdipendenti, ciascuna delle quali paragonabile agli anelli di una catena, che ne aumentano gradualmente il valore.



Figura 1.4: Le principali fasi della catena del valore dei Big Data

Ciascuno di questi passaggi ha alcune caratteristiche speciali e vede il potenziale coinvolgimento di diversi stakeholder, inclusi individui, organizzazioni, imprese, istituzioni pubbliche. La maggior parte di questi stakeholder sono coinvolti solo in parti selezionate della catena del valore, mentre solo in pochi sono più integrati verticalmente, come nel caso

di imprese in grado di raccogliere dati sui consumatori, quindi di archivarli, aggregarli ed infine elaborarli e utilizzarli nel proprio modello di business.

Prima di considerare in modo più approfondito tali fasi, è necessario evidenziare che tutti gli anelli della catena del valore dei dati presentano alcune barriere all'ingresso, che possono essere tecnologiche, comportamentali o legali. L'altezza di tali barriere determina dunque la concorrenza per la raccolta, l'archiviazione, l'analisi nonché la concorrenza nei mercati dei prodotti o servizi basati sui dati. Le determinanti di tali vincoli saranno specificate, per ognuna delle fasi, nel capitolo successivo.

1.3.1 Raccolta

La prima fase della catena del valore corrisponde alla *raccolta* dei dati. Vale la pena sottolineare che oggi, nell'attuale contesto in cui gran parte delle attività economiche e sociali sono migrate su internet e in cui pressoché tutti i contenuti media sono resi disponibili in formato digitale, per la maggior parte dei modelli di business delle imprese operanti online riveste una particolare importanza la raccolta dei dati derivanti dalle attività online degli utenti.

Tecnicamente l'acquisizione dei dati generati dagli utenti presuppone l'utilizzo di una serie di tecnologie diverse dedicate al loro tracciamento. A tal proposito, infatti i fornitori di servizi online possono rintracciare gli utenti sul web e raccogliere i loro dati utilizzando: cookies, web beacons o browser fingerprinting (Figura 1.5).



Figura 1.5: Tecnologie di tracciamento sul web [11]

I *cookies* sono piccoli file di testo utilizzati come identificatori e memorizzati sui dispositivi individuali al fine di tracciarne l'utilizzo. Il loro funzionamento è il seguente: ogni qual volta un utente visita un sito, il server fornisce all'utente un cookie che serve come forma di identificazione, qualora poi l'utente in un secondo momento visiterà di nuovo il sito, le

informazioni contenute nel cookie saranno ritrasmesse al server, che potrà quindi adattare il contenuto della pagina web alle informazioni che aveva su quel particolare utente.

I cookie sono classificati come *cookie di prima parte*, il cui dominio appartiene al fornitore di servizi online, cioè al sito a cui si accede, o come *cookie di terza parte*, il cui dominio, invece, appartiene a un'entità separata diversa dal fornitore di servizi. Questi strumenti tengono traccia dei dettagli di accesso nonché delle preferenze personali su un sito web. In particolare, i cookie di terze parti sono utilizzati per tutti gli annunci di retargeting e per la pubblicità comportamentale. Aggiungendo tag a una pagina, gli inserzionisti possono monitorare un utente o il suo dispositivo su diversi siti web. Ciò aiuta a costruire un profilo dell'utente in base alle sue abitudini, in modo che i messaggi possano essere più mirati ai suoi interessi [12].

I *web beacons* rappresentano brevi linee di codice che consentono la visualizzazione di immagini spesso minuscole e invisibili su siti web ed e-mail. Questa tecnologia di tracciamento opera inviando al server una richiesta ogni qual volta che si scarica un sito web o un'e-mail in cui è contenuto. Così facendo, il proprietario del web beacon è in grado di raccogliere dati sull'utente come l'ora di accesso, la posizione o l'indirizzo IP. Come per i cookies, esistono sia *web beacon di prima parte*, il cui dominio cioè appartiene al proprietario del sito web o al provider di posta elettronica, sia *web beacon di terze parti* che si collocano su siti web di altri provider per raccogliere informazioni sulle sessioni di navigazione dell'utente anche al di fuori dell'utilizzo dei loro servizi.

Per superare poi i casi in cui gli utenti eliminano i cookies dal loro browser oppure utilizzano altri strumenti volti a disabilitare le tecnologie di tracciamento sono state sviluppate tecnologie di tracciamento alternative come il *browser fingerprinting*, progettato appunto per funzionare anche quando i cookies sono disabilitati e che consente di risalire all'identità di un utente o di un dispositivo specifico sulla base di dati quali le configurazioni del browser, le informazioni dell'intestazione HTTP, i font installati e i componenti aggiuntivi. Il fingerprinting può essere utilizzato per vari scopi, come la valutazione delle prestazioni di un sito web per un utente o l'adattamento dell'interfaccia utente in base al tipo di dispositivo [12].

Esaminate le principali tecnologie di raccolta dei dati, occorre prendere in considerazione i *meccanismi principali di raccolta*: i dati possono essere raccolti attraverso un contatto diretto con la persona o l'oggetto da cui i dati vengono raccolti, si parla in questo caso di **raccolta di dati di prima parte**, indirettamente, rintracciando gli utenti in Internet con una **raccolta di dati di terza parte**, oppure ancora **acquistandoli da intermediari** di dati. In particolare:

- La *raccolta di dati di prima parte* avviene ogni qual volta un'azienda raccoglie informazioni direttamente dai propri clienti/utenti come parte dell'utilizzo dei beni o servizi da essa offerti. Ad esempio, i dati di prima parte di Google sono i dati che Google raccoglie dagli utenti quando questi utilizzano il suo motore di ricerca, la sua posta

elettronica e altri servizi posseduti e forniti dall'azienda. I dati raccolti in tal modo sono quindi una combinazione di input e permessi di accesso ai dati esistenti, forniti direttamente dagli utenti stessi durante la fase di registrazione ad un servizio online e/o all'atto del suo utilizzo. I vari fornitori di servizi online possono infatti raccogliere dati relativi all'utente in occasione della sua registrazione al servizio, solitamente un indirizzo e-mail o un numero di cellulare per impostare un account, ma anche informazioni aggiuntive, come nome, data di nascita, sesso e informazioni sul pagamento. Oltre alla creazione di un account, l'utilizzo di un servizio dipende tipicamente dall'accettazione delle condizioni di privacy e dalla concessione di permessi di accesso a determinate funzioni del dispositivo di un utente. Attraverso questi permessi, e a meno che l'utente non li revochi, il fornitore di servizi online è così in grado di accedere a molteplici dati quali le foto nella galleria immagini del dispositivo, la posizione, l'elenco dei contatti in rubrica. Una volta poi che l'utente inizia ad utilizzare il servizio e ad interagire con esso, la quantità di dati forniti che un'azienda può raccogliere dipende dalle funzioni del servizio stesso. Tanta maggiore, infatti, in questi casi sarà la possibilità data agli utenti di generare, piuttosto che consumare, contenuti tanto di più saranno i dati raccolti, si pensi ad esempio alle funzionalità del servizio di social network di Facebook che consente agli utenti, tra le molte altre cose, di pubblicare foto, video, post e live-streams, così come "likes" a pagine di marchi, squadre sportive, personaggi famosi ecc. [12].

- La raccolta dei dati degli utenti che non hanno un rapporto diretto con il fornitore di servizi online può avvenire anche attraverso il tracciamento di terze parti, si parla in tal caso di **raccolta di dati di terza parte**, definito da Robertson (2020) come:

“...una pratica che permette a un tracker di raccogliere grandi quantità di dati personali da una varietà di fonti di prima parte nell'ambiente online e attraverso diversi dispositivi come smartphone, tablet e laptop/computer, costruendo in ultima analisi un profilo utente completo” [13].

I terzi possono accettare tale tracciamento come parte di accordi commerciali per ricevere servizi di analisi di siti web e servizi di ad serving, ad esempio, così come nell'utilizzo di API⁶ proprietarie [14].

A tal proposito si rileva che il tracking da parte di terzi è prevalente sia sui siti web che sulle applicazioni. Sul web, il 95% dei 10.000 siti web più popolari al mondo contengono almeno un tracker di terze parti, con Google che rappresenta il tracker di maggior successo, seguito da Facebook e Twitter [15]. Per quanto riguarda invece le applicazioni, risulta che oltre il 90% delle applicazioni gratuite presenti sul Google Play

6 Acronimo di Application Programming Interface. Si tratta sostanzialmente di un'interfaccia di programmazione che permette ai programmatori esterni di accedere a determinate funzioni. Al giorno d'oggi, diversi produttori di software, tra cui SAP, Amazon, Google e Facebook, mettono a disposizione le loro API per facilitare l'accesso dei programmatori a componenti software, utili, tra le altre cose, per inserire contenuti nei servizi web. Ad esempio, sulla piattaforma di Google Maps, gli sviluppatori possono accedere ad alcune API e integrarle nelle proprie app, nei programmi e nei siti web. Come nel caso dell'API JavaScript di Maps che consente ai proprietari delle pagine web di aggiungere facilmente mappe interattive nel loro sito, funzione particolarmente interessante per i negozi o i ristoranti.

Store contengono ugualmente tracker di terze parti. In questo caso a detenere il primato è sempre Google seguito da Microsoft (incluso LinkedIn), Facebook e Twitter [16].

Da queste cifre si può evincere che la capacità di tracciamento e di raccolta di dati è concentrata nelle mani di un piccolo gruppo di aziende. La situazione così delineata è quindi indicativa di una sorta di “*data-opoli*” in cui essenzialmente vi sono poche piattaforme che rappresentano alcuni nodi di potere dove convergono la quasi totalità dei dati raccolti attraverso il tracciamento di terzi [13]. Questi giganti tech, concedendo infatti in outsourcing i sistemi di tracciamento e le API proprietarie agli sviluppatori di siti web e/o di applicazioni, sono in grado in tal modo di rientrare nella disponibilità dei dati acquisiti da questi ultimi e di rafforzare così ulteriormente una posizione già di per sé privilegiata per l’acquisizione diretta dei dati vista l’estrema popolarità dei loro servizi e delle loro applicazioni [10].

- Infine, nel caso in cui non si disponesse di un proprio grande set di dati di prima parte o si cercasse di arricchire i dati esistenti, le imprese nella fase di raccolta dei dati potrebbero acquisirli da intermediari, i cosiddetti **data brokers**, definiti dalla FTC statunitense come “*società la cui attività principale è la raccolta di informazioni personali sui consumatori provenienti da una varietà di fonti e aggregando, analizzando e condividendo tali informazioni, o informazioni da esso derivate, per scopi come la commercializzazione di prodotti, la verifica dell’identità di una persona o rilevamento di frodi*” [17].

Tali intermediari, attraverso il reperimento e la combinazione di dati provenienti da svariate fonti sia online che offline (quali i registri dei tribunali, dei veicoli e delle proprietà), sono in grado di creare dettagliati profili utente i quali sono poi messi a disposizione delle aziende per scopi che includono il marketing, la pubblicità e i prodotti finanziari. Tuttavia, proprio per via del fatto che combinano dati offline e online provenienti da più fonti diverse ed è così molto difficile per un consumatore conoscere e determinare come un broker di dati abbia ottenuto i suoi dati, si tratta di pratiche poco trasparenti che destano non poche preoccupazioni circa la privacy e la protezione dei dati.

1.3.2 Archiviazione

Successivamente alla fase di raccolta vi è quella di *archiviazione* dei dati in forme recuperabili in modo da poterli trattare. In considerazione del volume crescente di dati che vengono acquisiti, per questa attività si rendono necessari grandi data center, costituiti da grandi cluster di computer, dotati di memorie capienti che consentano una facile scalabilità e tempi di accesso e di trasferimento rapidi. I dati possono essere archiviati su server interni o su servizi di cloud computing esterni.

La necessità di trattare grandi volumi e varietà di dati, sempre più spesso in tempo reale è tale da rendere l’utilizzo di sistemi tradizionali non più funzionale. Se per quasi quaranta anni infatti i Relational Database Management Systems (RDBMS o modelli relazionali) sono stati la principale e quasi unica soluzione di archiviazione, per gli scenari Big Data

presentano caratteristiche che li rendono inadatti. Di conseguenza, sono state progettate tecnologie NoSQL al fine di introdurre soluzioni di archiviazione basate su modelli che non richiedono uno schema fisso.

1.3.3 Analisi

La terza fase della catena del valore dei Big Data è rappresentata dall'attività di *analisi*, la quale, in virtù del fatto che i dati di per sé hanno poco valore ma lo acquisiscono quando sono organizzati e elaborati, riveste un ruolo centrale. Quale elemento chiave del valore aggiunto, tale passaggio prevede l'organizzazione, l'integrazione e l'elaborazione analitica dei dati grezzi al fine di dedurre informazioni suscettibili di essere utilizzate per finalità economiche. In sostanza, questa attività si occupa di trasformare le informazioni derivanti da grandi moli di dati, spesso non strutturati, in conoscenza e viene condotta attraverso l'utilizzo di tecniche analitiche e algoritmi sofisticati.

Ed è proprio l'"intelligenza" di tali tecniche, congiuntamente all'elevato volume e varietà dei dati, che sta facendo emergere un nuovo paradigma analitico, cosiddetto *data driven*, in cui i dati non solo concorrono a verificare ipotesi teoriche con tecniche statistiche, ma anche a esplorare nuovi scenari e nuove teorie, nonché, più in generale a scoprire nuova conoscenza attraverso gli algoritmi di intelligenza artificiale. Questi particolari algoritmi, che auto-apprendono e migliorano il proprio funzionamento sulla base dell'esperienza ottenuta in iterazioni successive, sono così alla base di un approccio per la generazione di conoscenza del tutto innovativo dal punto di vista scientifico, che riconosce appunto ai dati una rilevanza centrale nell'ambito di questo nuovo paradigma analitico.

Infatti, se è vero che certamente i dati non sono l'unico input per lo sviluppo di questi strumenti- sono necessari anche forza lavoro competente e creativa, infrastrutture di storage e di calcolo – si può altresì riconoscere che rappresentano l'origine stessa della loro evoluzione in virtù del fatto che gli algoritmi di analisi apprendono e migliorano le loro performance man mano che sono alimentati da un numero crescente di dati. Pertanto, è proprio la disponibilità di sempre maggiori e nuove fonti di questi ultimi che consente il progressivo miglioramento degli algoritmi impiegati e /o lo sviluppo di nuovi algoritmi.

In base alle osservazioni precedenti, occorre notare che, relativamente a questa fase della catena del valore, i colossi dell'economia digitale (quali Google, Apple, Facebook, Amazon, Microsoft) appaiono godere di un vantaggio proprio perché, oltre a disporre di enormi quantità di dati (di prima e terze parti) si distinguono per capacità di investimento e finanziare uniche per cui sono in grado di dotarsi di soluzioni computazionali efficienti e di risorse umane di eccellenza [10].

1.3.4 Utilizzo

L'ultimo anello della catena del valore dei Big Data è rappresentato dall'*utilizzo* della conoscenza estratta dai dati. Vi sono diverse forme di utilizzo, tra le principali:

- In primo luogo, i Big Data possono essere utilizzati per il **miglioramento di prodotti o servizi** di un'azienda attraverso effetti di apprendimento. Come esempio si può

considerare il caso dei motori di ricerca web in cui si può ragionevolmente presumere che un maggior numero di ricerche contribuisce a migliorare e perfezionare il motore di ricerca e l'implementazione del suo algoritmo di supporto. Questo può migliorare la qualità dei risultati di ricerca, che a sua volta può poi portare a un maggior numero di persone che utilizzano quel particolare servizio.

- In secondo luogo, i Big Data possono essere utilizzati per **sfruttare nuove opportunità di business** attraverso l'**offerta di prodotti e servizi innovativi**. Attraverso il riutilizzo dei dati raccolti nel contesto di un servizio per uno scopo diverso, le imprese possono fornire nuovi servizi basati su questi dati. Basti considerare, ad esempio, i dati di posizione e sulla mobilità dei singoli utenti generati dagli operatori di rete mobili e dai telefoni cellulari che sono utilizzati dai fornitori di servizi di navigazione per mostrare le condizioni di traffico stradale e indirizzare meglio i loro utenti.
- In terzo luogo, poiché i Big Data potrebbero consentire alle imprese di ottenere una conoscenza altamente dettagliata dei bisogni e delle preferenze dei singoli consumatori, potrebbero anche essere utilizzati per sviluppare **modelli di business più orientati al target** indirizzando meglio i potenziali clienti e fornendo loro pubblicità, servizi o prodotti personalizzati. A tal proposito può servire da esempio la pubblicità online basata sul cosiddetto "behavioral targeting"⁷ (o anche "targeting comportamentale") di cui se ne parlerà più in dettaglio nella sezione 2.2.3 del capitolo successivo.

1.4 L'ecosistema dei Big Data

La raccolta, l'elaborazione e la conversione in valore dei Big Data avvengono in un complesso ecosistema caratterizzato da un intricato intreccio di interazioni in cui sono coinvolti diversi tipi di attori, come rappresentato nella Figura 1.6 [18].

Al centro dell'ecosistema dei Big Data, dove si osservano molti dei problemi di concorrenza che saranno in seguito discussi, le *piattaforme* operano come principale interfaccia tra i consumatori e gli altri attori del mercato. In linea di principio, se ne possono distinguere due categorie principali: le *piattaforme di attenzione* e le *piattaforme di matching*.

- Le *piattaforme di attenzione*, come i motori di ricerca o i social network, forniscono tipicamente un insieme di servizi "gratuiti" che sono so ^{Data} dalla pubblicità venduta su base "per-click". In questo modo, i consumatori, invece di pagare un prezzo monetario per il servizio, pagano con la loro ^{Money} attenzione, ad esempio dovendo visionare dei contenuti intervallati da pubblicità o essendo obbligati a guardare una pubblicità prima di accedere ad un contenuto video. Oltre che con la loro attenzione, i consumatori pagano anche con i loro dati, i quali sono poi utilizzati per migliorare la qualità dei

⁷ Per "targeting comportamentale" si intende il servizio di pubblicità online a specifici utenti sulla base di profili (completi) degli utenti generati dall'osservazione delle loro abitudini di navigazione.

servizi e per indirizzare meglio la pubblicità, permettendo alla piattaforma di attrarre nuovi consumatori e di far pagare agli inserzionisti un costo più alto per click.

- Le *piattaforme di matching*, invece, forniscono un mercato in cui diversi tipi di attori possono interagire, come acquirenti e venditori, datori di lavoro e dipendenti, o anche singoli individui nei siti di incontri online. Le piattaforme di matching guadagnano denaro applicando tariffe fisse per l'accesso alla piattaforma e tariffe variabili per transazione. Spesso, il gruppo di utenti con una maggiore elasticità della domanda è sovvenzionato dall'altro gruppo (per esempio, i clienti non pagano per utilizzare i siti di shopping; le persone in cerca di lavoro non pagano per utilizzare i siti di lavoro, e così via). Anche in questo caso, gli utenti della piattaforma forniscono dati che vengono utilizzati per migliorare la qualità del servizio e degli algoritmi di matching, portando in ultima analisi a un maggior numero di transazioni.

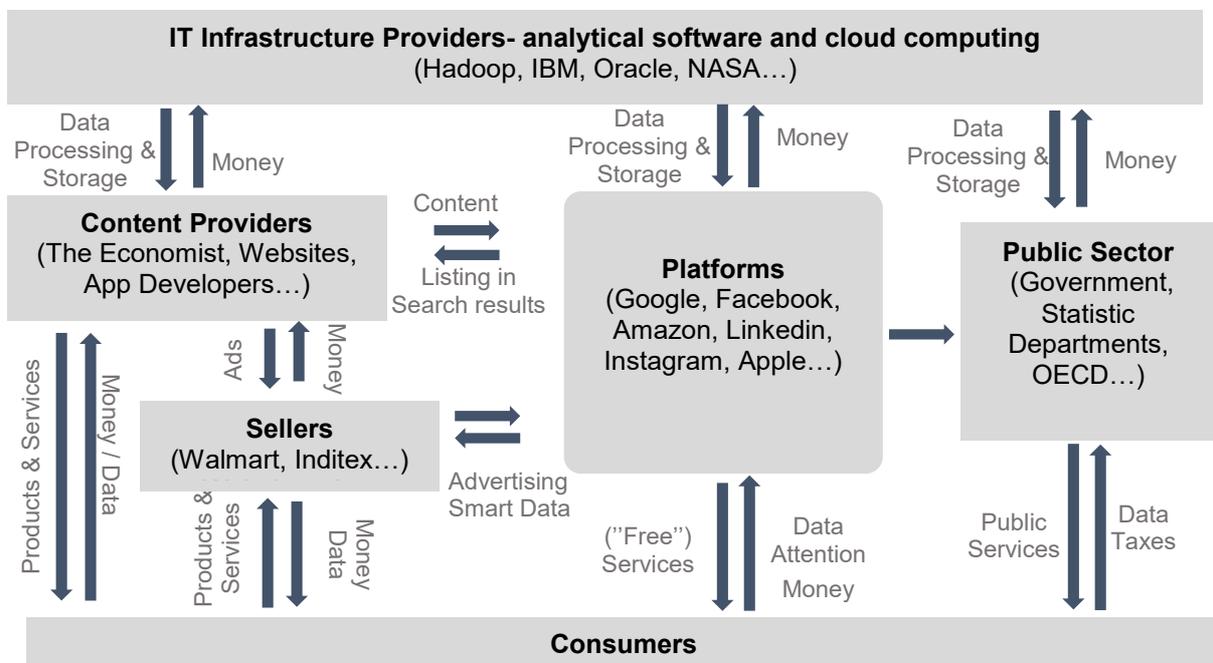


Figura 1.6: Ecosistema dei Big Data [18]

Un altro insieme di attori dell'ecosistema dei Big Data sono i *fornitori di contenuti* come giornali, siti web e sviluppatori di applicazioni che creano i contenuti informativi disponibili in molte piattaforme, in cambio di una posizione nell'elenco dei risultati. I contenuti creati non vengono visualizzati solo dai motori di ricerca come parte del loro core business, ma anche da altre piattaforme come i social network, che hanno bisogno di contenuti creativi per attirare e trattenere l'attenzione dei consumatori e mantenere alti i livelli di traffico. I fornitori di contenuti guadagnano denaro sia vendendo il loro prodotto direttamente ai consumatori, sia vendendo spazi pubblicitari ai venditori. Tuttavia, poiché non dispongono dei Big Data necessari per effettuare un'adeguata pubblicità mirata, è sempre più frequente

che i siti web pubblichino, ad esempio, annunci pubblicitari attraverso la piattaforma (come quella di Google) e che ricevano una quota dei ricavi pubblicitari.

I principali sovvenzionatori degli attori discussi finora sono i **venditori**, che offrono prodotti e servizi al consumatore finale in cambio di denaro. Tra questi vi sono produttori, grossisti, professionisti, agenzie immobiliari, consulenti, istituzioni finanziarie e qualsiasi altro tipo di imprese che possono utilizzare i canali di marketing delle piattaforme per convincere i consumatori ad acquistare i loro prodotti. Vi è un certo numero di grandi venditori che possono crescere abbastanza in scala da poter utilizzare i Big Data da soli, proprio come nel caso di Amazon, Tesco o Target (il secondo più grande discount degli Stati Uniti). Raccogliendo i dati delle transazioni online, delle carte fedeltà e dei moduli presentati dai consumatori, a volte in cambio di sconti sui prezzi e di prodotti gratuiti, queste aziende sono in grado di aumentare ulteriormente le loro dimensioni e di acquisire una distanza irrecuperabile dai concorrenti più piccoli, che non hanno la scala e le costose infrastrutture per elaborare i Big Data e diventare validi concorrenti.

Le operazioni condotte dagli utenti dei Big Data sono supportate in modo determinante dai **fornitori di infrastrutture informatiche (IT)**, come Hadoop, IBM e Oracle. In effetti, le aziende che si occupano di innovazione dei dati si trovano rapidamente a dover affrontare petabyte di dati che sono costosi da memorizzare e ancora più difficili da elaborare, per i quali mancano le risorse. I fornitori di infrastrutture IT non solo sviluppano il software adeguato a gestire i Big Data, ma soprattutto forniscono il cloud computing e lo storage, ovvero agiscono come data center di terze parti dove le aziende possono archiviare ed elaborare i loro dati on-demand. Questi data center sono solitamente costituiti da grandi cluster di computer collegati da reti locali veloci, costantemente operativi e che beneficiano in larga misura di economie di scala.

Infine, sul lato opposto dell'ecosistema si trova il **settore pubblico**, comprendente il governo centrale e locale, così come gli ospedali pubblici, le cliniche, gli enti di previdenza sociale e tanti altri enti. Il settore pubblico rappresenta uno dei settori dell'economia a maggiore intensità di dati che, raccoglie Big Data dai cittadini e, occasionalmente, dalle piattaforme e dai venditori, quando questi ultimi sono tenuti a fornire informazioni per rispettare la legge. I dati così raccolti sono utilizzati per migliorare i prodotti e i servizi offerti, nonché per sostenere la ricerca scientifica.

È importante sottolineare che il settore pubblico non costituisce solo un attore attivo nell'universo dei Big Data, ma è anche uno stimolatore e un modellatore dell'ecosistema: le autorità di regolamentazione insieme ai governi infatti svolgono un ruolo fondamentale nell'influenzare la dimensione e la forma dell'ecosistema, nonché il valore da esso creato. Da un lato, le autorità di regolamentazione hanno il mandato di proteggere la sicurezza dei dati e i diritti alla privacy dei cittadini. Pertanto, essi cercano di proteggere i consumatori dal potenziale abuso della loro identità. Dall'altro lato, le autorità di regolamentazione bilanciano questo mandato con la necessità di promuovere la crescita economica e il benessere dei cittadini [18].

1.5 Categorie di dati

Qualunque discussione sui *Big Data* e sulle loro implicazioni in termini di concorrenza e di privacy dovrebbe includere un riconoscimento delle molte diverse forme di dati. Una categorizzazione, infatti, si rende necessaria per poter individuare successivamente le tipologie di dati che sono rilevanti per la valutazione della concorrenza ma la cui raccolta e utilizzo al contempo sollevano non poche preoccupazioni ai sensi della legge sulla privacy e sulla protezione dei dati.

Gli approcci alla categorizzazione dei dati sono diversi. In primo luogo, i dati possono essere classificati in base ai diversi tipi di informazioni che forniscono e che a loro volta possono essere di diversi gradi di utilità e di disponibilità per una data impresa.

I dati ad esempio possono fornire informazioni su *individui*, sulle loro preferenze, sui loro comportamenti, la loro posizione geografica, ecc.; possono quindi fornire informazioni su *entità economiche* in riferimento al fatturato realizzato e al numero di transazioni commerciali effettuate; oppure ancora possono offrire informazioni su *oggetti* fornendo ad esempio la posizione o la velocità di un'automobile [19].

Oggetto di questa analisi saranno in particolare i dati sugli individui, denominati da qui in avanti anche come consumatori, in quanto la raccolta e lo sfruttamento di tali dati, vista la pervasività di internet nelle vite quotidiane di ciascuno di noi, costituiscono il core business di molte delle aziende più redditizie al mondo e quindi un campo particolarmente rilevante nell'economia dei Big Data. I dati dei consumatori sono eterogenei e il valore privato e sociale dei dati dipende da come vengono utilizzati e può variare per i diversi attori della catena del valore [20]. Pertanto, la comprensione delle varie dimensioni dei dati sui consumatori è importante per comprendere il valore potenziale dei diversi tipi di dati sui consumatori.

A questo proposito, i dati sui consumatori possono essere classificati: (i) in base all'origine dei dati, (ii) a seconda che i dati sui consumatori possano essere personalmente identificabili, (iii) in base alla loro natura strutturata o meno. Ciascuna di queste classificazioni sarà discussa nelle sezioni seguenti.

1.5.1 Classificazione in base all'origine dei dati

I dati dei consumatori possono essere classificati in base alla loro origine. A tal proposito Abrams (2014) [21] ha proposto le seguenti quattro categorie (Figura 1.7): dati *volontari*, inseriti direttamente dall'utente, *osservati* e quindi registrati mentre l'utente utilizza un servizio, *derivati* e infine *dedotti* attraverso l'analisi predittiva.



Figura 1.7: Tipi di dati in base al modo in cui sono stati generati

Dati volontari

Si tratta di tutti quei dati forniti attivamente dall'utente. Sono da considerarsi tali dunque i dati che un consumatore inserisce all'atto della registrazione su un social network oppure ad una applicazione di messaggistica istantanea o a qualunque servizio online che richiede delle credenziali di accesso, oppure ancora i dati creati ogni qualvolta l'utente interagisce con il servizio ad esempio quando effettua un acquisto online ed inserisce i dati relativi alla sua carta di credito e il suo indirizzo per la consegna. Rientrano in questa categoria anche i dati volontari quali post, foto, video e storie postati dall'utente sui social network.

In sostanza le diverse tipologie di dati volontari sono tutte caratterizzate da un denominatore comune: il consumatore, anche se non sempre ne considera le implicazioni, è pienamente consapevole delle azioni che sta intraprendendo e partecipa attivamente alla creazione dei dati.

Dati osservati

A differenza dei dati forniti volontariamente, i dati osservati rappresentano i dati raccolti automaticamente mentre l'utente utilizza il servizio. Le informazioni direttamente correlate al comportamento e all'attività degli utenti durante l'uso dei servizi includono, ad esempio, informazioni sul tempo trascorso utilizzando un'applicazione di messaggistica, su ciò che l'utente cerca online o il tempo in cui un utente interrompe la riproduzione di un particolare contenuto video su un servizio di condivisione video. Altre informazioni sono di natura più indiretta o "passiva", come le informazioni sulla localizzazione, le caratteristiche del dispositivo e del software, l'indirizzo IP ed in generale tutte le attività di navigazione online. Oltre a ciò, altri dati raccolti automaticamente possono includere dati provenienti da cookie, web beacon e altre tecnologie pubblicitarie (discussi nella sezione 1.3.1).

Le due categorie di dati sopra descritte solitamente non sono in grado di per sé di fornire informazioni di valore per un'impresa, essi necessitano di essere combinati ed elaborati per creare dei profili utente. Si pensi, ad esempio ad un singolo dato riguardante un video che un utente ha guardato, tale dato da solo potrebbe non fornire alcuna informazione di valore per il fornitore del servizio. A tal scopo dunque, quest'ultimo dovrà creare i propri *dati derivati*, come il numero aggregato di visualizzazioni di un video per utente o un segmento

di pubblico o *dedurre dati* sull'utente come la sua preferenza per un certo tipo di contenuto video

Dati derivati

Sono quindi dati sull'utente che vengono derivati da un'analisi abbastanza meccanica di altri dati che questi ha fornito o che l'impresa ha osservato durante l'utilizzo del servizio. Questa tipologia di dati può includere, ad esempio, la preferenza di un utente per i documentari in base alla loro storia di visione, oppure la spesa media per visita di un utente su un sito di e-commerce oppure ancora la percentuale di articoli restituiti rispetto a quelli acquistati. Ciascuno dei nuovi prodotti di calcolo è un elemento di dati che potrebbe essere utilizzato da un'organizzazione per comprendere meglio il comportamento o prendere decisioni che riguardano il consumatore, il quale generalmente non è consapevole della creazione di questi dati.

Dati dedotti

Differentemente dai dati derivati, i quali tendono ad essere basati su semplici calcoli, i dati dedotti vengono creati attraverso analisi più complesse e non banali dei dati volontari e osservati, condotte con tecniche sofisticate, le quali permettono appunto di dedurre nuove informazioni utilizzando dati già esistenti. Ad esempio, un sito di e-commerce di moda potrebbe analizzare i singoli prodotti che un utente ha visionato per dedurre, sia pure con un certo margine di errore, il sesso del visitatore. Oppure ancora un esempio di dati dedotti può essere una valutazione del credito calcolata sulla base di una serie di fattori rilevanti per la storia finanziaria di un individuo. In sostanza, i dati dedotti fanno uso dei dati forniti e osservati per trovare modelli e fare ipotesi sulle preferenze o gli attributi degli utenti, e per questo motivo sono utilizzati per lo più a sostegno della pubblicità online.

Un'altra classificazione, sempre legata all'origine dei dati, prevede di aggiungere un'ulteriore categoria a quelle sopra definite- i dati *acquisiti da terzi* - e di organizzarle infine in due gruppi a seconda che i dati siano stati *raccolti* o *creati* dall'impresa, come mostrato nella Figura 1.8 [12].

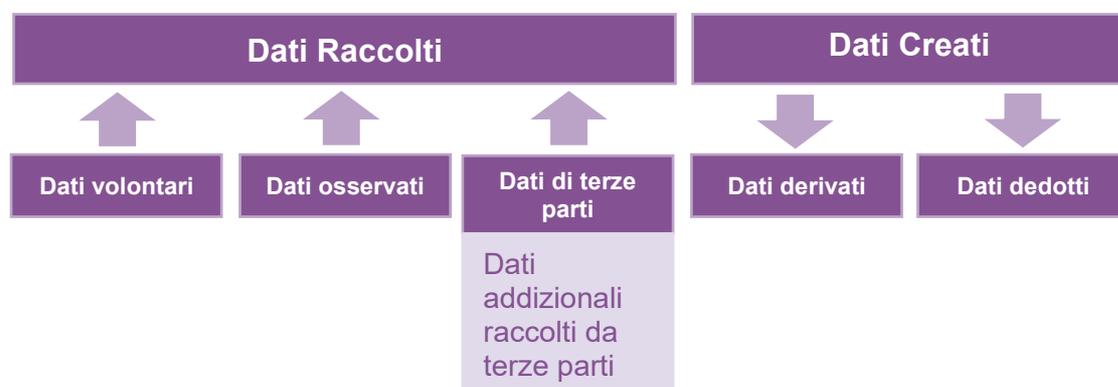


Figura 1.8: Tassonomia dei dati raccolti e creati da un fornitore di servizi online

Dati acquisiti da terzi parti

Si tratta dei dati che un'azienda raccoglie, come risultato di un acquisto da un intermediario o di una partnership commerciale che prevede l'utilizzo di tecniche di tracciamento, da fonti terze e che quindi sono ottenuti al di fuori delle interazioni dirette con i suoi utenti.

Come rappresentato in Figura 1.9, l'importanza e il valore che i dati di terzi rivestono per un'azienda dipendono strettamente dal suo modello di business e sono quindi legati al fatto che questa tragga i suoi ricavi principalmente dalla pubblicità o dagli abbonamenti.

Laddove infatti un'azienda basi il suo modello di ricavi principalmente sulla pubblicità online, cercherà con maggior probabilità di raccogliere informazioni sugli interessi e il comportamento online dell'utente anche al di fuori dell'utilizzo del suo servizio, derivandone così un profilo più ampio e accurato e aumentando, di conseguenza, l'attrattività del servizio di advertising per gli inserzionisti. Al contrario, i servizi basati prevalentemente su abbonamenti faranno meno ricorso a fonti terze per i dati e utilizzeranno i dati generati dagli utenti durante l'utilizzo dei loro servizi per migliorare i loro prodotti e guidare i miglioramenti aziendali interni.



Figura 1.9: L'impatto dei modelli di ricavo sull'uso dei dati [12]

Ad esempio, Google e Microsoft adottano approcci ai dati molto diversi. Il fatturato di Microsoft si basa principalmente sulla vendita di prodotti e servizi, per cui Microsoft rappresenta più un elaboratore di dati, piuttosto che un controllore, e usa i dati per migliorare i suoi servizi, piuttosto che trattarli come un fattore chiave di reddito. Allo stesso modo, il servizio di abbonamento Netflix non ha quasi alcun interesse per i dati demografici in quanto li ha trovati poco utili per la programmazione o la personalizzazione e preferisce lavorare con i dati sulle preferenze rivelate [12]. Al contrario, attori come Google e Facebook sono interessati ad una raccolta più ampia di dati che va oltre quella che avviene attraverso i loro servizi. Questo approccio gli consente di sviluppare profili utente più dettagliati che possono mettere a disposizione degli inserzionisti per il targeting degli annunci pubblicitari.

Infine, è importante sottolineare che la classificazione basata sull'origine dei dati ha una duplice rilevanza politica. *In primo luogo*, il modo in cui i dati hanno origine riflette i diversi livelli di consapevolezza e di controllo che gli interessati possono avere sui dati raccolti su o da loro. Ed è chiaro che la consapevolezza dei consumatori costituisce un aspetto importante per affrontare i problemi di informazione asimmetrica associati alla raccolta e all'uso dei loro dati. Questi diversi livelli di consapevolezza sono in gran parte riconducibili al grado in cui l'utente può aver fornito i dati in moto attivo o passivo.

In particolare, infatti, i consumatori saranno in genere più consapevoli delle informazioni che hanno fornito volontariamente, potrebbero avere una certa consapevolezza dei dati osservati, mentre potrebbero avere molto meno consapevolezza, o addirittura nessuna, dei dati derivati e/o acquisiti da terzi⁸. *In secondo luogo*, i dati volontari e i dati osservati sono probabilmente meno soggetti, rispetto ai dati derivati e acquisiti, alle affermazioni dell'azienda secondo cui i dati sono stati "creati" da essa stessa. Ciò è rilevante per gli incentivi delle aziende a condividere questi vari tipi di dati, nonché per l'effetto delle politiche relative alla portabilità e all'interoperabilità dei dati, che potrebbero dover bilanciare i compromessi tra lo stimolo della concorrenza, la protezione degli incentivi agli investimenti e i problemi di privacy di terzi [14].

1.5.2 Classificazione in base al grado di identificazione personale

I dati dei consumatori possono essere classificati anche in base al grado di identificazione personale. Innanzitutto, occorre precisare che con il termine **dati personali** ci si riferisce a "qualsiasi informazione relativa ad un individuo identificato o identificabile" [22]. Di conseguenza, sono da intendersi tali tutti dati che consentono l'identificazione diretta o indiretta di un individuo attraverso un qualunque identificatore.

Poiché un grado più elevato di identificabilità sarebbe tipicamente associato a rischi relativi alla privacy più elevati, l'uso o il trattamento di questi dati è subordinato al rispetto di condizioni specificate dalle politiche sulla privacy.

In base a questa classificazione, si possono distinguere quattro categorie di dati (Figura 1.10) corrispondenti a quattro diversi stati di identificabilità personale, elencati di seguito secondo un ordine decrescente del rischio di protezione dei dati associato [23]



Figura 1.10: Tipi di dati dei consumatori personali e non personali

⁸ In parte, ciò è dovuto al fatto che gli utenti hanno un basso livello di consapevolezza delle politiche sulla privacy le quali generalmente sono troppo lunghe e di difficile lettura per il consumatore medio.

- *Dati identificativi*: dati che possono essere associati in modo univoco ad una persona specifica perché ne consentono l'identificazione diretta attraverso un identificatore come un nome, un numero identificativo, dati relativi all'ubicazione, un identificatore online o simili. Appartengono a questa categoria una particolare tipologia di dati, i *dati sensibili*, i quali comprendono informazioni rivelanti l'origine razziale o etnica, opinioni politiche, credenze religiose o filosofiche, o appartenenza sindacale, o dati genetici, dati biometrici, dati relativi alla vita sessuale o all'orientamento sessuale di una persona fisica, ecc.
- *Dati pseudonimizzati*: dati per i quali tutti gli identificatori sono stati sostituiti da pseudonimi in modo tale che nessun dato possa essere attribuibile ad un individuo specifico interessato senza l'utilizzo di informazioni aggiuntive, come ad esempio una chiave che fa la corrispondenza tra un codice e gli identificativi comuni dell'individuo. Poiché gli pseudonimi applicati possono essere comunque invertiti dall'operatore che li ha assegnati, che può dunque ristabilire il collegamento, sono da considerarsi a tutti gli effetti come dati personali.
- *Dati anonimizzati*: sono dati personali resi anonimi in modo tale che la persona specifica interessata non sia più identificabile.
- *Dati anonimi*: Dati i cui attributi sono alterati (ad esempio, i valori degli attributi sono randomizzati o generalizzati) in modo tale che vi sia un ragionevole grado di fiducia che una persona non possa essere identificata, direttamente o indirettamente, dai soli dati o in combinazione con altri dati, tenendo conto di tutti i mezzi che possono essere ragionevolmente utilizzati a tal scopo dal responsabile del trattamento o da qualsiasi altra persona. Sono da considerarsi tali, ad esempio, i dati statistici che non contengono voci a livello individuale e che sono combinati a partire da informazioni su un numero sufficiente di persone diverse che gli attributi a livello individuale non sono identificabili.

La rilevanza di questa categorizzazione è che, laddove i dati non possano essere attribuiti ad un individuo, è più probabile che questi siano in grado di fornire vantaggi economici e competitivi senza far scattare preoccupazioni in materia di privacy. Tuttavia, la natura dinamica dei dati personali pone degli interrogativi sulla capacità di salvaguardare i dati anonimi da una nuova identificazione. Gli attuali sviluppi nell'analisi dei dati e nell'intelligenza artificiale hanno reso, infatti, più facile la ricombinazione di informazioni che di per sé non sono personali ma che, in combinazione, possono consentire l'identificazione di uno specifico individuo, tanto è che *“Una volta che qualsiasi dato è stato collegato all'identità reale di una persona, qualsiasi associazione tra questi dati e un'identità virtuale rompe l'anonimato di quest'ultima”* [24]. Ciò in un certo senso offusca la distinzione tra dati personali e non personali e sfida qualsiasi approccio normativo che determini l'applicabilità di diritti, restrizioni e obblighi sulla sola base di un concetto statico di "dati personali"[25].

1.5.3 Classificazione in base alla struttura dei dati

Infine, i dati possono essere pure differenziati in base al fatto che siano o meno “strutturati”. Questa peculiarità ha delle conseguenze sulle possibilità di estrarre valore economico dai dati nonché sulle modalità necessarie per farlo. In particolare:

- I **dati strutturati** sono tutti quei dati che possono essere organizzati in modo ordinato sulla base di un modello che definisce una serie di campi, il tipo di dati che questi contengono e come si relazionano tra di loro. Un esempio di dati strutturati sarebbe una banca dati degli indirizzi dei consumatori che contiene informazioni relative al nome, cognome, indirizzo, età, numeri di telefono, ecc. di ciascuno di essi. Proprio per via della loro struttura ordinata possono essere elaborati e utilizzati più facilmente per scopi commerciali rispetto ai dati non strutturati, almeno se si considerano i metodi tradizionali.
- I **dati non strutturati**, al contrario, non sono conformi ad un modello specifico e proprio per via del fatto che non hanno un formato o un'organizzazione predefinita rendono la loro analisi più complicata, motivo per il quale devono essere trattati con algoritmi diversi, più innovativi e più recenti affinché acquisiscano un determinato valore commerciale. I dati non strutturati possono comprendere audio, video, foto, messaggi di testo, e-mail.

Le due categorie presentano una serie di differenze, elencate in modo sintetico nella Figura 1.11.

Dati Strutturati	Dati non strutturati
✓ Difficili da raccogliere	✓ Facili da raccogliere
✓ Convenienti da elaborare	✓ Più costosi da elaborare
✓ Intuizioni limitate	✓ Intuizioni quasi infinite
✓ Richiedono una partecipazione attiva	✓ Richiedono solo la presenza
✓ La trasparenza promuove la privacy	✓ Mancanza di trasparenza, privacy

Figura 1.11: *Principali differenze tra i dati strutturati e non strutturati*

In primo luogo, i dati strutturati sono difficili da raccogliere ma convenienti da elaborare con i metodi tradizionali. Infatti, se si pensa ai questionari, che sono un modo comune di raccogliere dati strutturati, questi richiedono comunque che una persona sia attenta e disposta ad impegnarsi, per cui devono essere progettati secondo una lunghezza tale da convincere il soggetto a portare a compimento la sua compilazione. Per tali motivi, la

quantità di dati che possono essere estratti in questo modo, o attraverso qualsiasi altro metodo che faccia ricorso a dei moduli predefiniti, non è affatto paragonabile alle migliaia di elementi che possono essere estratti dai dati non strutturati. Al contrario, invece, i dati non strutturati sono più facili da raccogliere ma la loro elaborazione è più costosa a causa proprio della loro natura disorganizzata che rende necessarie tecniche più complesse per estrarre le informazioni. Per la raccolta di dati non strutturati, infatti, non è necessario che ci siano un questionario appositamente progettato e l'attenzione immediata dell'utente, occorre solo la sua presenza e la sua interazione con il servizio online. Per cui, data soprattutto la costosa e troppo limitata nel tempo attenzione dell'utente, è indubbio il vantaggio che ne deriva relativamente al potenziale volume di dati raccolti.

In secondo luogo, anche se la raccolta di dati strutturati può essere difficile, ha però i suoi vantaggi: la natura stessa dei dati strutturati, infatti, rende palese agli individui le informazioni che l'entità interessata ai loro dati vuole sapere da loro e consente di conseguenza al consumatore la scelta delle informazioni di cui rendere l'altra parte al corrente. I confini tra le parti sono, dunque, in tal modo fissati e più espliciti all'interno di questi moduli strutturati. Ad esempio, nell'ipotesi in cui un'azienda conduca un'indagine strutturata nella quale si chiedano informazioni sulla vita coniugale di una persona, su quanto spesso questa si arrabbi con i figli e altre informazioni sulla sua vita privata, una persona può ritenere che queste domande non siano affatto di competenza dell'azienda e di conseguenza decidere di non compilare quel modulo non permettendo così all'azienda di ottenere quei dati. Diverso è il caso, invece, dei dati non strutturati in quanto l'utente non è pienamente consapevole di cosa acconsente quando utilizza un servizio attraverso il quale può generare dati di questo tipo. Ancor di più per via del fatto che i dati non strutturati che si forniscono potrebbero essere utilizzati per scopi completamente diversi da quelli ipotizzati. Le aziende potrebbero combinare dati non strutturati in un modo che esula le loro competenze. A tal proposito, ha destato non poco sconcerto negli Stati Uniti il caso di un signore che ha accusato Target di adottare un comportamento irresponsabile per avere inviato alla figlia, ancora adolescente, coupon e volantini su prodotti premaman e per neonati, salvo poi scoprire che la figlia fosse incinta e che Target lo avesse capito prima di lui [24].

Questa storia è esplicativa di come, per ragioni di business, le aziende potrebbero esercitare una certa invadenza nella sfera privata individuale e pone quindi non poche preoccupazioni sulla privacy, soprattutto in ragione di due motivi: (i) innanzitutto perché un unico dato non strutturato può essere utilizzato in molti modi diversi, tanto è che la parte che raccoglie i dati può non definire in anticipo le informazioni da estrarre da questi e iniziare a raccogliere dati quali file audio, immagini, video in modo casuale senza alcuno scopo e capire poi in un secondo momento per cosa utilizzarli, (ii) secondariamente perché le aziende hanno imparato ad applicare un certo tatto nell'impiego di tali dati - a differenza di quanto avvenuto nel caso Target - e ad utilizzarli in un modo in cui l'individuo non ne ha quasi neppure la piena consapevolezza. Ad esempio, un utente potrebbe aver accettato che il rilevatore di volto sul suo smartphone raccolga, quali dati non strutturati, i dati relativi al

suo volto nella convinzione che questi sarebbero stati usati esclusivamente per consentire di bloccare e sbloccare in modo sicuro il telefono e quindi di ricevere un servizio migliore. Tuttavia, gli stessi dati potrebbero essere utilizzati in modi leggermente diversi dal primo scopo dichiarato. L'azienda, infatti, potrebbe impiegarli in casi d'uso che destano meno preoccupazioni circa la privacy, come consentire all'utente di creare un avatar sul suo telefono che mostri il suo sorriso o il suo cipiglio in tempo reale, così come potrebbero essere impiegati in casi d'uso più personali, ad esempio l'azienda potrebbe prendere quegli stessi dati facciali e venderli per informare gli inserzionisti su come i loro annunci vengono recepiti da un utente [25].

In conclusione, appare evidente che le diverse tipologie di dati descritte in questo paragrafo pongono gradi diversi di preoccupazione in riferimento alla privacy e alla protezione dei dati. A tal proposito, nel capitolo 4 saranno considerate le indicazioni e i regolamenti che la normativa europea propone come risposta ai diversi problemi di privacy che la raccolta e l'utilizzo dei vari tipi di dati del consumatore pongono.

CAPITOLO 2

Big Data: fonte di potere di mercato?

I Big Data sono diventati uno dei temi più importanti nel vivace dibattito in corso sulla concorrenza e la regolamentazione nell'economia digitale. La raccolta e l'uso dei dati da parte delle aziende a fini commerciali non è un fenomeno nuovo, tuttavia la situazione attuale, come osservato nel capitolo precedente, si distingue per la scala e la portata dei dati raccolti ed elaborati nonché per la loro estrema importanza per molti dei modelli di business delle aziende tecnologiche di maggior successo. Proprio perché i dati sugli utenti e sulle loro preferenze costituiscono un input sempre più essenziale per l'attività delle imprese online è normale chiedersi se e in quali condizioni il possesso e l'utilizzo dei Big Data possa conferire potere di mercato e quindi un vantaggio competitivo duraturo.

In particolare, il ruolo dei dati nel processo competitivo tra le aziende nell'economia digitale e soprattutto tra le piattaforme online è una questione controversa che ha iniziato ad attirare l'attenzione dei policy maker e degli studiosi della comunità del diritto della concorrenza. Infatti, sebbene da alcuni anni i Big Data siano considerati uno dei più promettenti motori dello sviluppo economico, più recentemente costituiscono fonte di preoccupazione per il possibile potere di mercato che può essere accumulato da parte di chi raccoglie grandi quantità di dati e ha la capacità di elaborarli.

Ad alimentare questa preoccupazione vi è l'evidenza del successo commerciale dirompente di una manciata di aziende in grado di raccogliere enormi quantità di dati dai consumatori e di offrire loro servizi basati su essi. La figura 2.1 conferma l'ascesa, in diversi mercati digitali, di aziende superstar per le quali l'estrazione di conoscenza da grandi volumi di dati rappresenta un input fondamentale. Si può osservare che tali aziende controllano la quota di mercato più grande nei principali servizi forniti online.

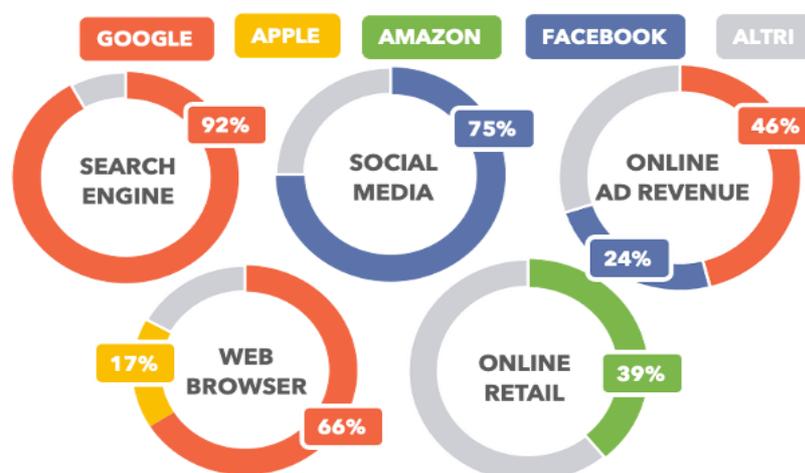


Figura 2.1: Quote di mercato globali per società (2019). Fonte: GSStatCounter, eMarkete, Statista

Per tale motivo, vi sono crescenti preoccupazioni che questa tendenza sia il segno di uno spostamento dei mercati digitali verso una dinamica del tipo "winner-take-all", dove le aziende che raggiungono la leadership di mercato, o riescono a controllarne una quota significativa, possono ribaltare il mercato a loro favore, diventando molto difficili da "detronizzare" e creando una serie di problemi nell'area della concorrenza [26].

In generale non vi è un accordo sul ruolo dei Big Data come base di un potere di mercato duraturo e di conseguenza non vi è un accordo sul ruolo del diritto della concorrenza nei confronti del business dei Big Data. I commentatori sono generalmente divisi in due campi: uno a favore di un'applicazione più proattiva dell'antitrust nel regno dei Big Data, e uno contrario a tale intervento. Da un lato si sostiene infatti che i Big Data non costituiscono una base per un potere di mercato: ad esempio Tucker C. (2019) e Lerner (2014) sostengono che i Big Data non creano una barriera significativa all'ingresso e che quindi è improbabile che siano una fonte di potere di mercato. Tale affermazione, si basa tra l'altro, sulla natura non esclusiva e non rivale⁹ dei dati e sulla presunta facilità di raccolta degli stessi, senza tener conto di molte potenziali barriere all'ingresso [27] [28]. Dall'altro lato, si sostiene che dati superiori possono non solo portare *temporaneamente* a una posizione di mercato dominante, ma che - a causa della maggiore interazione con l'utente che deriva dal servizio migliore - consentono all'azienda dominante di migliorare la qualità della sua base dati più velocemente rispetto ai potenziali concorrenti e possono quindi condurre a un *vantaggio permanente*. A tal proposito ad esempio il Garante Europeo della Protezione dei Dati ha dichiarato che "la raccolta e il controllo di enormi quantità di dati personali sono una fonte di potere di mercato per i maggiori operatori del mercato globale dei servizi Internet" [29]. Altri studiosi, tra cui Stucke e Grunes, concentrandosi principalmente sulla raccolta dei dati e sugli effetti di rete dei Big Data concludono che "l'economia dei dati...favorisce la concentrazione e la dominanza del mercato" [30].

Questo capitolo cercherà di offrire una migliore comprensione dei Big Data come base (potenziale) per il potere di mercato, evidenziando come le aziende possono trasformare i dati in valore e quali fattori possono promuovere il potere di mercato (persistente). La discussione sarà condotta in tre step in cui si discuterà di:

1. Come le imprese possono stabilire un potere dei dati ("*Data Power*") attraverso la raccolta e l'elaborazione di enormi moli di dati che consentono la creazione di profili utente dettagliati.
2. Come le informazioni sulle caratteristiche, preferenze e il comportamento degli utenti, derivanti dai loro dati, possono essere usati per generare valore economico.

⁹ I dati sono considerati beni "non rivali" nel senso che la raccolta e l'utilizzo da parte di un provider non pregiudica la raccolta e l'utilizzo da parte di un altro provider.

3. Quali caratteristiche, fattori e condotte delle imprese possono favorire la concentrazione del potere di mercato nel contesto delle imprese con potere di dati.

2.1 Stabilire un “Data Power”

Come già osservato nel capitolo precedente, la raccolta e l’elaborazione dei dati degli utenti costituiscono un’attività essenziale alla base dei modelli di business di tutte le aziende online. I dati degli utenti sono un input chiave al punto tale che le aziende molto spesso offrono ai singoli individui servizi gratuiti con l’obiettivo di attrarne l’attenzione e di acquisirne informazioni. Da un lato i singoli utenti forniscono essi stessi dei dati volontariamente sotto forma, ad esempio, di informazioni sul profilo, foto, liste di amici o contatti sui social network e ricerche effettuate sui motori di ricerca o sulle piattaforme di e-commerce. Dall’altro lato, i fornitori di servizi online osservano, deducono e derivano dati attraverso l’analisi del comportamento e delle abitudini degli utenti. Tali dati sono poi ricombinati per formare dei profili comportamentali i quali vengono sfruttati commercialmente per vendere pubblicità “mirata”, prodotti, servizi personalizzati e/o per rivenderli ad altre aziende.

Nonostante però le attività di raccolta, analisi e utilizzo dei dati siano attività comuni a quasi tutte le aziende online, non tutte godono di un “*Data Power*” sostanziale, ovvero di una significativa capacità di raccolta ed elaborazione dei dati e di creazione di profili utente completi e dettagliati. *In primo luogo*, perché solo le imprese già affermate con una base utenti già sviluppata ottengono più dati in input: le aziende più piccole o i nuovi operatori infatti raccolgono di norma meno dati di prima parte rispetto alle controparti più grandi e consolidate. Inoltre, oltre ai dati di prima parte, le aziende possono anche utilizzare “dati di terze parti”, ovvero raccolti seguendo l’attività dell’utente su altri servizi online. Anche in questo caso solo poche imprese hanno una significativa capacità di creare set di dati più ampi non limitati ai soli propri clienti (vedi sezione 1.3.1). *In secondo luogo*, un “*Data Power*” potrebbe anche essere conferito dall’enorme varietà dei dati di cui beneficiano solo alcune imprese che fungono attualmente da importanti gateway verso il mondo digitale, ovvero che possono attingere ai dati da diverse fonti in quanto hanno costituito un ecosistema capace di offrire un’ampia gamma di servizi per imprese e consumatori. *In terzo luogo*, solo alcune imprese sono capaci di analizzare i dati in tempo reale ed utilizzare dati storici per migliorare le applicazioni o gli algoritmi esistenti e svilupparne di nuovi. Tali miglioramenti o creazioni consentono in altre parole in un certo senso di capitalizzare il valore transitorio dei dati in un valore più permanente, soprattutto quando si utilizzano dati che diventano rapidamente obsoleti.

In definitiva, in tale contesto, coloro che godono di un maggior numero di “obò” per raccogliere dati, che dispongono di un consistente database con cui confrontare i nuovi dati, o che possiedono strumenti unici per la loro sintesi e la loro analisi, riescono a stabilire un significativo *Data Power* sul quale poi basare un vantaggio competitivo. Diverse sono però le barriere all’ingresso come evidenziato di seguito.

2.1.1 Barriere all'ingresso ai mercati dei Big Data

A causa della presunta natura onnipresente e non rivale delle informazioni, è stato sostenuto che non è plausibile che un'impresa ottenga un potere per quanto riguarda i dati degli utenti da cui potrebbe derivarne potenzialmente un potere di mercato [31]. Tuttavia, anche se la raccolta, l'analisi e l'utilizzo dei Big Data siano un aspetto cruciale della concorrenza per la maggior parte dei fornitori di servizi online, ciò non significa che tutti siano liberi di farlo e soprattutto che siano liberi di farlo alle stesse condizioni degli operatori già affermati. Barriere all'ingresso e all'espansione infatti potrebbero incidere sulla concorrenza a lungo termine, limitando la crescita delle imprese più piccole e l'ingresso di nuove imprese.

Rubinfeld e Gal (2017) [31] hanno intrapreso un'analisi approfondita della catena del valore dei Big Data per individuare le possibili barriere all'ingresso associate alla raccolta, all'archiviazione, all'analisi e all'utilizzo dei dati che impongono ad alcuni operatori del mercato di trovare un percorso alternativo e meno efficiente rispetto a quello degli operatori storici. Queste sono riassunte nella Tabella 2.1 e discusse più dettagliatamente di seguito.

Tabella 2.2: Barriere all'ingresso nella supply chain dei dati [14]

	Barriere tecniche	Barriere Legali	Barriere comportamentali
Raccolta	<ul style="list-style-type: none"> • Unicità dei dati, o accesso a essi • Lato offerta: economie di scala, di scopo, learning by doing • Lato domanda: effetti di rete, mercati bilaterali 	<ul style="list-style-type: none"> • Leggi sulla protezione dei dati e sulla privacy • Proprietà dei dati 	<ul style="list-style-type: none"> • Accordi di esclusiva • Prezzi di accesso e condizioni • Bias comportamentali
Archiviazione	<ul style="list-style-type: none"> • Costi di storage 	<ul style="list-style-type: none"> • Leggi sulla protezione dei dati e sulla privacy 	<ul style="list-style-type: none"> • Lock-in e switching costs
Analisi	<ul style="list-style-type: none"> • Assenza di interoperabilità • Strumenti analitici 		
Uso	<ul style="list-style-type: none"> • Impossibilità di localizzare e raggiungere consumatori rilevanti • Assenza di interoperabilità 	<ul style="list-style-type: none"> • Leggi sulla protezione dei dati e sulla privacy • Leggi antidiscriminazione 	<ul style="list-style-type: none"> • Limitazioni contrattuali

Per quanto riguarda la **raccolta** dei dati, Rubinfeld e Gal (2017) hanno rilevato che se un'azienda avesse accesso a dati unici, questo potrebbe essere difficile da ricreare senza incorrere in sostanziali costi irrecuperabili. Sebbene alcuni dati, infatti, possono essere raccolti in modo semplice ed economico in parallelo da molte aziende (come ad esempio la posizione di un utente che può essere osservata contemporaneamente da molte applicazioni di smartphone a costi molto bassi), altri come ad esempio l'accesso ai post e alle reazioni dei consumatori su un social network dominante potrebbero essere difficili e costosi da ricreare. A questo proposito, l'analisi dovrebbe considerare l'accesso dell'azienda sia ai dati di prima parte che a quelli di terzi. Potrebbero pure sorgere barriere tecnologiche sul lato dell'offerta laddove gli operatori storici, sostenendo costi per lo più fissi e bassi costi marginali, abbiano realizzato notevoli economie di scala o di scopo, o del tipo "learning by

doing". Sul lato della domanda, invece, la presenza di effetti di rete, di mercati bilaterali e il fatto che per raccogliere alcuni tipi di dati sia necessario entrare in un mercato correlato (quello che Rubinfeld e Gal chiamano ingresso a due livelli), può anche aumentare i costi irre recuperabili necessari per entrare nel mercato rilevante. Inoltre, barriere comportamentali possono essere erette o direttamente dalle aziende, come gli accordi di accesso esclusivo a una fonte unica di dati e i prezzi o le condizioni di accesso discriminatorie, oppure dagli stessi utenti che preferiscono, anche a servizi potenzialmente migliori, i servizi offerti dal first-mover solo per questioni di fiducia, reputazione e riconoscimento del marchio (si parla in tal caso di pregiudizi o bias comportamentali).

Per quanto riguarda l'**archiviazione** occorre dire che i progressi tecnologici (come lo sviluppo del cloud computing) hanno ridotto in modo significativo le barriere all'ingresso nei mercati dei Big Data, il che rende più improbabile che aspetti tecnici costituiscano una barriera all'ingresso di quanto non lo fossero in precedenza. Probabilmente il maggiore ostacolo alla concorrenza nel campo dell'archiviazione dei dati è rappresentato dai costi di lock-in e di switching oltre che dalle barriere legali, come le limitazioni su dove possono essere conservati (fisicamente) i dati personali di qualcuno.

Anche le fasi di **sintesi e analisi** dei dati, presentano una serie di potenziali barriere tecniche, tra cui i costi associati allo sviluppo di strumenti analitici e i limiti all'interoperabilità dei dati. Ogni titolare dei dati, infatti, li organizza in modo che si adattino alle proprie esigenze e preferenze. Tale organizzazione può creare però una barriera per la sintesi con altri dati o per il loro utilizzo da parte di altri.

Prendendo in considerazione, poi, le potenziali barriere associate all'**uso** dei dati, si rilevano le difficoltà nel localizzare e raggiungere i consumatori rilevanti, ad esempio nel caso in cui la piattaforma che i consumatori utilizzano è controllata da uno dei suoi rivali, nonché le limitazioni contrattuali alla portabilità dei dati dell'utente. Tali restrizioni limitano la capacità dell'utente di esportare i dati personali da un'applicazione all'altra senza incontrare troppe difficoltà, ne aumentano quindi i costi di switching e generano effetti lock-in.

Inoltre, barriere legali come la protezione dei dati e le leggi sulla privacy possono aumentare le barriere all'ingresso relativamente sia all'attività di raccolta dei dati che a quelle di archiviazione e utilizzo. L'altezza di queste barriere sarà influenzata, tra l'altro, dall'ambito di applicazione della legislazione e dal fatto che questa incorpori o meno norme sulla responsabilità. Un'altra serie di ostacoli legali deriva da problemi di proprietà dei dati, che incidono sulla facilità di accesso ai dati nonostante la loro natura non rivale. Si potrebbe temere che l'onere finanziario imposto alle piccole imprese o alle nuove imprese a seguito del rispetto di tali norme possa scoraggiarle dall'entrare nel mercato, riducendo la pressione concorrenziale e limitando il potenziale di nuovi prodotti e servizi innovativi.

Occorre, infine, dire che le barriere all'ingresso non sono invariabilmente basse o, al contrario, alte nei mercati online. La realtà è che la loro entità è da analizzare caso per caso ed è specifica per ciascun mercato.

2.2 Uso dei Big Data e Vantaggio Competitivo

Di per sé la quantità, la varietà e la qualità dei dati in possesso di un'azienda, sebbene siano alcune delle determinanti della sua forza competitiva, non costituiscono un indicatore adeguato del suo vantaggio competitivo, tantomeno di un potere di mercato duraturo.

Innanzitutto, affinché un'azienda possa godere di un *vantaggio competitivo data-driven* deve essere in grado di monetizzare¹⁰ efficacemente i profili utenti creati dall'analisi dei dati. Questi, infatti, possono essere sfruttati commercialmente per vendere pubblicità mirata, prodotti e servizi personalizzati e/o per rivenderli ad altre aziende. Il fatturato generato da un fornitore online attraverso la monetizzazione dei dati tramite i modi appena citati può costituire un'indicazione della sua forza competitiva.

In questa sezione si analizzeranno alcuni modi attraverso i quali le informazioni sulle caratteristiche, preferenze e comportamenti degli utenti derivati dai loro dati possono essere usati per generare valore economico.

2.2.1 Servizi personalizzati

Per servizi personalizzati si intende la personalizzazione dei servizi e delle esperienze di acquisto secondo le preferenze individuali dei consumatori. Poiché la personalizzazione viene sempre più percepita come una fonte di vantaggio competitivo, vi è una tendenza crescente verso il perseguimento di strategie di personalizzazione abilitate dalla tecnologia.

Le aziende online utilizzano tecnologie di personalizzazione avanzate come il filtraggio dei contenuti o l'analisi dei flussi di clic per personalizzare le loro offerte web in base alle preferenze individuali degli utenti o ai comportamenti d'uso passati. Per di più oggi la personalizzazione è sempre più possibile grazie all'interattività offerta dal Web. Infatti, negli ultimi anni, attraverso la proliferazione di tecnologie di interfaccia utente più sofisticate, gli utenti sono sempre più incoraggiati a partecipare al processo di personalizzazione. Basti pensare ai pulsanti "Pollice su/giù" o "Mi piace/non mi piace" che consentono agli utenti di rivelare le loro preferenze relativamente agli attributi di un prodotto o servizio. Le conoscenze estratte dalle "tracce" che i clienti lasciano durante la navigazione su un sito web possono poi essere integrate senza soluzione di continuità con algoritmi e software per personalizzare automaticamente i contenuti per i singoli individui.

¹⁰ Nel commercio online, per "monetizzazione" si intende la capacità di un provider di generare ricavi dai contenuti, dai servizi o dai prodotti offerti agli utenti, che spesso sono forniti gratuitamente.

Un singolo cliente online può incontrare diversi casi di personalizzazione. In primo luogo, un'azienda può offrire una personalizzazione basata su caratteristiche specifiche individuali come il nome, l'indirizzo di spedizione e la modalità di consegna preferita, le preferenze sugli sconti per i volumi, ecc. L'azienda può anche offrire servizi quali gli avvisi istantanei inviati al cliente quando i prezzi di particolari prodotti della sua "watch list" scendono a un livello predefinito o quando un'asta si avvicina alla chiusura. Inoltre, le aziende sul Web possono anche personalizzare l'esperienza di acquisto del cliente fornendo raccomandazioni personalizzate. Alcuni servizi online consentono addirittura agli utenti di personalizzare altri attributi come i colori delle loro pagine web, il nome con cui un individuo preferisce essere accolto e la classifica dei risultati di ricerca in base alle storie di acquisto e di ricerca del cliente [33].

La personalizzazione dei servizi ha il potenziale di massimizzare le opportunità di business immediate e future di un'azienda online attraverso varie dinamiche. Innanzitutto, a differenza della sua controparte offline, la personalizzazione online di per sé è raramente utilizzata per estrarre i premi di prezzo. Ad esempio, sia Amazon.com che Barnes&Noble.com offrono servizi di personalizzazione simili, come i promemoria e le opzioni di spedizione personalizzate, mentre nessuno di essi fa pagare esplicitamente i consumatori per questi servizi. Tuttavia, quando i clienti forniscono informazioni quali i loro anniversari, Amazon.com può utilizzare queste informazioni per indirizzare i clienti verso prodotti correlati e offrire loro specifici incentivi all'acquisto [33].

Tale strategia ha il preciso scopo, attraverso la capacità di rivolgersi ai clienti in modo individuale, di:

- Aumentare la loro soddisfazione e fidelizzazione, scoraggiando di conseguenza il passaggio ad altri fornitori ed aumentando i costi di switching. Ad esempio, un consumatore che ha già personalizzato le informazioni relative alle spedizioni e ai regali su un rivenditore basato sul web può mostrare una certa inerzia nel passare ad un concorrente anche se ci possono essere alcune differenze marginali nei prezzi, in quanto tale passaggio implicherebbe di nuovo numerosi moduli di compilazione.
- Aumentare anche la "stickiness" (viscosità) di un sito web, ovvero aumentare la permanenza degli utenti sul sito ad assistere ai contenuti web. Ciò dunque comporta un aumento dei ricavi pubblicitari oltre che un aumento della willingness to pay degli utenti [34].

A questo proposito, la personalizzazione del web è stata riconosciuta come una strategia di differenziazione chiave che aiuta le aziende ad **attrarre e mantenere i clienti** e, in ultima analisi, a rendere i loro modelli di business più sostenibili attraverso un aumento dei flussi di cassa e una maggiore redditività [34].

Occorre infine dire che poiché la personalizzazione dipende in modo critico, oltre che dalla capacità delle aziende di acquisire ed elaborare le informazioni dei consumatori, dalla disponibilità degli stessi a condividerle, gli investimenti nella personalizzazione online potrebbero essere gravemente compromessi se gli utenti non utilizzassero questi servizi

perché non disposti ad essere profilati, ovvero per motivi di privacy. Affinché dunque i consumatori online rivelino le loro preferenze è necessario che il beneficio da loro quantificato dei servizi personalizzati superi la perdita di privacy delle informazioni.

2.2.2 Raccomandazioni personalizzate

Internet e le tecnologie correlate hanno permesso alle aziende di offrire ai clienti una varietà di opzioni molto maggiore. Gli assortimenti enormi, tuttavia, sono vantaggiosi per i consumatori solo se le tecnologie di ricerca li supportano nella navigazione attraverso l'enorme varietà di prodotti, aiutandoli a identificare quelli che si adattano alle loro preferenze. Infatti, man mano che le aziende online ampliano l'offerta aumentano anche la quantità di informazioni che i clienti devono elaborare prima di poter scegliere quali articoli soddisfano i loro bisogni. Una soluzione a questo problema di sovraccarico di informazioni è l'uso di sistemi di raccomandazione, ovvero algoritmi capaci, sulla base di un'analisi del comportamento d'acquisto passato del cliente come previsione per quello futuro, di suggerire tra i prodotti del catalogo quelli che potrebbero soddisfarlo maggiormente, riducendo così i costi di ricerca per prendere decisioni di acquisto.

Generalmente si possono distinguere, in base alle modalità di formulazione delle raccomandazioni, due tipi di sistemi di raccomandazioni personalizzate [35]:

- *Content-based recommender systems* (sistemi di raccomandazione basati sul contenuto): in cui all'utente sono raccomandati elementi simili a quelli che ha preferito in passato o prodotti con proprietà simili. È il caso di uno degli strumenti utilizzati da Amazon.com che raccomanda l'acquisto di un pacchetto di due prodotti.
- *Collaborative recommender systems* (sistemi di raccomandazione collaborativi): in cui all'utente sono consigliati oggetti che persone con gusti e preferenze simili hanno acquistato in passato; ad esempio la funzione di Amazon.com con la dicitura "I clienti che hanno acquistato questo articolo hanno anche comprato" raccomanda gli altri prodotti acquistati frequentemente dai clienti che hanno già acquistato il prodotto selezionato.

Per quanto riguarda le ricadute in termini di redditività delle aziende che adottano questi sistemi, Adomavicius et al (2017) hanno esplorato il loro l'impatto sul comportamento dei consumatori dal punto di vista economico o decisionale, dimostrando che le raccomandazioni online aumentano in modo significativo la disponibilità del consumatore a pagare nella direzione della raccomandazione [36].

Inoltre, questi sistemi contribuiscono ad aumentare le vendite attraverso il cross-selling (vendita incrociata), ovvero attraverso il suggerimento al cliente di prodotti aggiuntivi da acquistare. Per di più, grazie alla riduzione dei costi di ricerca, consentono ai clienti di spostare la loro domanda dai blockbuster verso nicchie che meglio si adattano alle loro preferenze. Ne deriva così un aumento del surplus di consumo e di profitti, a suggerire che

le aziende online possono utilizzare queste tecnologie come base per un vantaggio competitivo [37].

2.2.3 Targeted advertising

Un'altra attività che consente alle imprese online di generare maggiori profitti attraverso la raccolta e l'analisi dei Big Data è la pubblicità mirata (*targeted advertising*). Poiché, come già osservato, i dati sul comportamento online dei consumatori (i siti che visitano, i video che guardano, gli articoli letti, ciò che cercano su un motore di ricerca ecc.) e le tecnologie di tracciamento consentono di sviluppare profili utente altamente dettagliati, questi possono essere utilizzati per indirizzare annunci ai consumatori che sono mirati individualmente in base ai loro interessi e preferenze. Come semplice esempio si consideri una rete pubblicitaria che tiene traccia delle visite dei consumatori ai siti web: se un consumatore visita diversi siti web di automobili, la rete presumerà un suo interesse a riguardo e gli proporrà annunci pertinenti. Di conseguenza, quando due persone visitano uno stesso sito Web contemporaneamente, una potrebbe visualizzare annunci di automobili mentre l'altra, che aveva visitato siti web di mobili, potrebbe visualizzare annunci di mobili [38].

Esistono diverse forme di pubblicità online mirata. Le principali sono [23]:

- La **pubblicità di ricerca** o **search advertising** permette alle aziende di pagare per far sì che le loro inserzioni siano incluse o classificate come prioritarie tra i risultati restituiti in risposta alla query di ricerca di un consumatore. Naturalmente si tratta di una forma di pubblicità mirata in quanto fornisce alle aziende l'accesso diretto ai consumatori nel momento esatto in cui stanno cercando informazioni circa uno specifico prodotto o servizio.
- La **pubblicità non di ricerca** o **non-search advertising**, la cui forma principale è la **pubblicità display**. Si tratta di banner pubblicitari che compaiono sullo schermo digitale e il cui contenuto può includere testo, immagini, video, ecc. Questa forma di pubblicità mirata sta diventando sempre più importante nel mercato della pubblicità online e poiché, soprattutto in relazione alla pubblicità di ricerca, prevede un ecosistema relativamente complesso è utile presentare in modo più preciso i diversi intermediari coinvolti in esso.

Innanzitutto, occorre precisare che ogni operazione che ha ad oggetto una pubblicità display coinvolge almeno:

- un **Editore** o **Publisher** che vende opportunità pubblicitarie ("inventario pubblicitario") e pubblica l'annuncio al suo pubblico, ad esempio, sul suo sito web, social network, app, podcast; L'editore inserisce annunci mirati propri o di terze parti sulla base delle informazioni che ha sui consumatori.

- un **Inserzionista** o **Advertiser** che acquista l'opportunità di pubblicizzare il proprio prodotto o servizio ai consumatori attraverso un annuncio pubblicato sul sito dell'editore.

Tuttavia, come illustrato di seguito in figura 2.2 la pubblicità display può assumere diverse forme [23]:

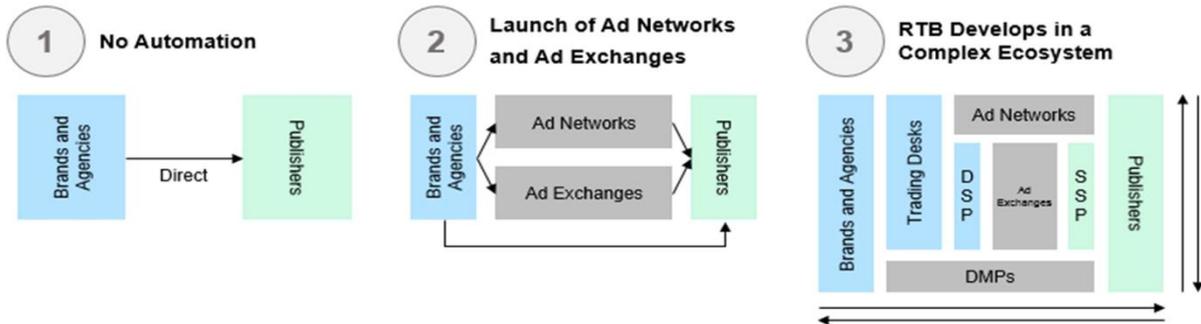


Figura 2.2: Forme diverse di targeted advertising [23]

- 1) *Nessuna Automazione*. Rappresenta la forma più semplice, in cui l'Advertiser si rivolge direttamente al Publisher (magari tramite un'agenzia pubblicitaria) per acquistare uno spazio pubblicitario. Solitamente questa forma è più comune nella pubblicità sui media tradizionali. Molto spesso, invece, per la pubblicità online l'operazione pubblicitaria è condotta con l'assistenza di un certo numero di fornitori di tecnologia pubblicitaria di terze parti che mirano a facilitare l'acquisto e/o la vendita di inventario di annunci, il targeting più fine dei consumatori, e la misurazione del comportamento del consumatore a seguito della pubblicità, come evidente nei successivi due casi.
- 2) *Reti pubblicitarie e Scambi di annunci*. Per vendere il proprio spazio pubblicitario l'editore può anche stipulare un contratto con una rete pubblicitaria o uno scambio di annunci. In particolare, le *reti pubblicitarie* fungono da intermediari tra gli inserzionisti che sono alla ricerca di un pubblico per promuovere i propri prodotti/servizi e gli editori, ovvero siti web e social network che vogliono ospitare pubblicità online. Tra gli esempi vi sono DoubleClick e AdMob di proprietà di Google, Audience Network di proprietà di Facebook, ecc. Uno scambio di annunci è invece una piattaforma automatizzata dalla quale viene venduto un inventario da più reti pubblicitarie. Tra gli esempi Double Click e Microsoft Ad Exchange.
- 3) *Offerte in tempo reale* (RTB=real time bidding). Nei sistemi RTB, gli scambi di annunci fungono da intermediari tra domanda e offerta di spazi pubblicitari, in particolare forniscono un mercato virtuale dove gli editori mettono all'asta, in tempo reale, il loro spazio pubblicitario agli inserzionisti idonei. In questa operazione sono coinvolte anche le piattaforme del lato dell'offerta (Supply Side Platform, SSP), che aiutano gli editori a vendere i loro spazi pubblicitari, le piattaforme del lato della domanda (Demand Side Platforms, DSP), che danno accesso agli acquirenti inserzionisti e le piattaforme di gestione dei dati (Data Management Platforms, DMP) che vendono i dati sugli utenti agli altri players. Nel dettaglio le RTB funzionano nel modo seguente: ogni volta che un

utente si collega al sito web di un editore, la piattaforma di gestione dei dati associata ne identifica il profilo utente attraverso i suoi cookie e lo invia allo scambio di annunci, il quale a sua volta lo fornisce alle piattaforme del lato della domanda (DMP). A questo punto segue un'asta in due fasi: nella prima, ogni DSP avvia un'asta con i suoi inserzionisti sulla base dello specifico profilo utente fornito; i vincitori della prima asta partecipano quindi ad una seconda presso lo scambio di annunci. Ai vincitori di quest'ultima, in conclusione, sarà concessa l'opportunità di far visualizzare i propri annunci sulla pagina web consultata dall'utente [23].

Impatto della pubblicità mirata per i consumatori, gli inserzionisti e gli editori

Innanzitutto, poiché questa forma di pubblicità prevede che gli annunci siano mirati sulla base degli interessi e delle preferenze del consumatore, essa comporta un'alta probabilità di un miglior "abbinamento" tra questo tipo di messaggi pubblicitari e gli interessi dell'utente e tende quindi a migliorare l'efficacia dell'annuncio. Ne segue che ciò può aumentare il benessere dei consumatori introducendoli a prodotti e servizi più pertinenti e riducendo i loro costi di ricerca.

In particolare, inoltre, per il motivo di cui sopra, gli inserzionisti si aspettano di sperimentare per gli annunci mirati tassi di click e tassi di conversione più elevati rispetto agli annunci generici online [39]. Non solo, la pubblicità mirata ha il potenziale di ridurre i messaggi inutili perché si rivolge solo ai clienti interessati, il che riduce i costi di pubblicità per gli inserzionisti e stimola conseguentemente la domanda di pubblicità online. Gli editori beneficiano in tal modo di maggiori ricavi, i quali possono poi essere parzialmente trasferiti ai consumatori in termini di minori costi di abbonamento o di maggiore qualità del servizio. Tuttavia, non si può neppure escludere che gli annunci mirati potrebbero suscitare nei consumatori sentimenti di invadenza e preoccupazioni per la privacy. D'altra parte, però, per via della sua maggior efficacia rispetto alla regular advertising il fastidio del consumatore, se esposto ad un numero inferiore di annunci più pertinenti rispetto al caso della pubblicità normale, potrebbe essere addirittura minore.

In conclusione, è attraverso queste attività che le imprese possono utilizzare le informazioni sulle caratteristiche, preferenze e comportamento degli utenti derivanti dai loro dati per generare valore economico e passare dunque da un Data Power a un vantaggio competitivo come si osserva nella figura 2.3 seguente.



Figura 2.3: Dal Data Power al Vantaggio Competitivo

2.3 Fattori che facilitano il potere di mercato

In ultima analisi, una volta individuate le possibili fonti di vantaggio competitivo, occorre chiedersi quali sono i fattori principali che possono favorire la concentrazione del potere di mercato nel contesto delle imprese con Data Power.

Fattori specifici dei mercati basati sui dati, quali le elevate economie di scala e di scopo, gli effetti di rete, i cicli di feedback, in determinate condizioni, possono conferire alle imprese un vantaggio competitivo duraturo, che può aiutarle a deviare in modo persistente la concorrenza attuale e futura. Questi elementi, infatti, potrebbero rendere il potere di mercato “viscoso” e di conseguenza debole la minaccia che un concorrente possa spostare l'operatore in carica.

2.3.1 Economie di scala e di scopo

Certamente le economie di scala e di scopo rappresentano dei fattori che hanno una relazione positiva con la concentrazione del mercato, per cui devono essere presi in considerazione nella valutazione del potere delle imprese ad alta intensità di dati.

Molti mercati digitali sono caratterizzati da alti costi fissi e bassi costi variabili: i motori di ricerca, ad esempio, necessitano di alti livelli di investimento iniziale per creare gli algoritmi e i centri di calcolo necessari per elaborare i dati ma, una volta sviluppati, i costi per un utente aggiuntivo sono piuttosto bassi. Di conseguenza, poiché i costi incrementali per l'offerta di unità aggiuntive diminuiscono con l'aumentare del numero di clienti serviti, le imprese digitali mostrano *ritorni di scala* molto forti.

Inoltre, le aziende data-driven possono beneficiare di risparmi sui costi dovuti, oltre che alle loro dimensioni, alla loro presenza in una gamma di servizi diversi, traendo così vantaggio dalle *economie di scopo*. Utilizzando, infatti, le relazioni esistenti con i clienti e i fornitori, la condivisione delle competenze tecniche, la condivisione e la fusione dei dati, un'azienda può ridurre i costi o migliorare la qualità del servizio sviluppando ecosistemi ed espandendo la propria presenza su una gamma di servizi correlati.

Tuttavia, poiché si sostiene che i vantaggi relativi alla disponibilità di grandi moli di dati siano soggetti a rendimenti decrescenti, ovvero che il valore di avere informazioni aggiuntive diminuisce con l'aumentare della quantità di dati [28], il potenziale di questi fattori nello stabilire un potere di mercato dipenderà quindi dal volume al quale i rendimenti dei dati aggiuntivi iniziano a diminuire. Se i benefici derivanti dalla disponibilità di informazioni aggiuntive iniziano a diminuire solo in presenza di una quantità di dati molto elevata, grandi volumi di dati possono dar luogo, rispettivamente, a un vantaggio competitivo per le piattaforme già presenti sul mercato e a una barriera all'ingresso per i nuovi operatori e potenziali concorrenti.

2.3.2 Feedback loops

Un altro meccanismo che può favorire un potere di mercato persistente è il cosiddetto “*feedback loop*”, che consente di attirare gli utenti attratti da prodotti e servizi di qualità superiore e più mirati, creando in tal modo un circolo virtuoso che rende difficile per i piccoli operatori sfidare le grandi aziende. Questo meccanismo potrebbe condurre naturalmente ad un indebolimento della concorrenza indipendentemente dal fatto che vi sia o meno un comportamento anticoncorrenziale delle aziende già affermate. Tanto è vero che l’OCSE osserva che:

"la piattaforma dominante può non fare nulla che possa essere adeguatamente qualificato come anticoncorrenziale, eppure il ciclo di feedback può rafforzare la posizione dominante e impedire alle piattaforme rivali di acquisire clienti" [18].

In particolare, si distinguono due diversi cicli di feedback:

Il **ciclo di feedback degli utenti** presuppone che un’azienda online che gode di una base utenti più ampia e che è quindi capace di raccogliere più dati, può beneficiare di una migliore comprensione dei consumatori e delle loro esigenze attraverso il loro utilizzo per migliorare la qualità del servizio erogato e attirare a sua volta ancora più utenti. La figura 2.4 illustra le diverse fasi del ciclo in questione.



Figura 2.4: Il ciclo di feedback degli utenti [28]

Il **ciclo di feedback di monetizzazione** presuppone, invece, che un’azienda online con una base utenti più ampia, che è quindi in grado di raccogliere più dati, è più capace di indirizzare gli annunci pubblicitari e conseguentemente di monetizzare meglio il proprio servizio. La migliore targetizzazione, infatti, produce maggiori ricavi in quanto da un lato aumenta la probabilità che gli utenti facciano clic sugli annunci (remunerati comunemente sulla base del modello pay-per-click) che visualizzano, dall’altro lato attrae altri inserzionisti. Queste maggiori entrate possono, poi, essere investite per migliorare ulteriormente la qualità del servizio, attirando così un maggior numero di utenti, il che di per sé contribuisce ad attrarre gli inserzionisti per via di effetti di rete indiretti. Le diverse fasi di questo ciclo sono illustrate nella figura 2.5.

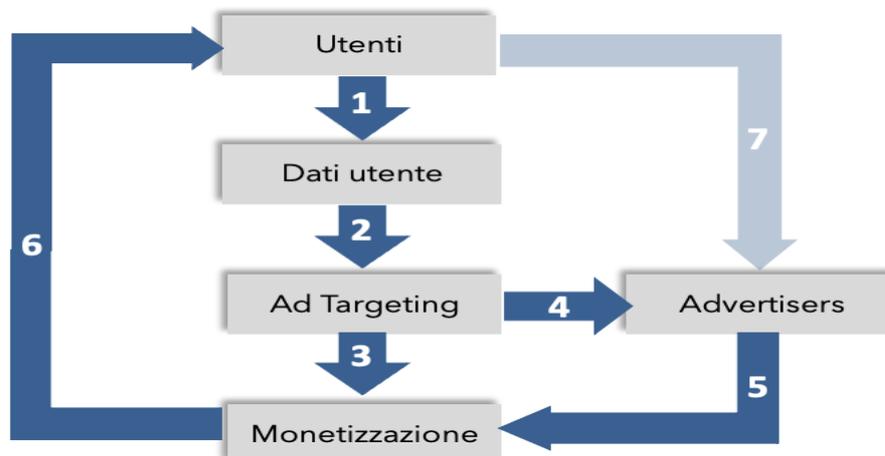


Figura 2.5: Il ciclo di feedback di monetizzazione [28]

Se si verificano questi cicli di feedback è chiaro che una piccola azienda online con meno utenti e meno dati avrà ancora più difficoltà ad attrarre nuovi utenti e a monetizzare efficacemente il proprio servizio. Mancando, infatti, di una quantità di dati comparabili a quelli di un'azienda già affermata, la sua capacità di migliorare la qualità e di attrarre un maggior numero di utenti sarà quantomeno inibita.

A titolo esemplificativo, se un motore di ricerca ha solo mille interrogazioni giornaliere, i suoi algoritmi hanno meno dati per imparare a restituire agli utenti risposte pertinenti e ricerche correlate. Con risultati di ricerca di qualità inferiore, sarà difficile attirare molti utenti dai motori di ricerca più grandi; con meno utenti, il motore di ricerca attirerà meno inserzionisti, il che significa meno occasioni per gli utenti di cliccare sui risultati di ricerca a pagamento e meno entrate pubblicitarie per espandere la piattaforma ad altri servizi [18].

Tuttavia, l'entità e l'esistenza di questi circuiti deve essere testata caso per caso e dipende in particolare, nel primo ciclo dalla relazione tra i dati e la qualità del servizio, che a sua volta dipende dal tipo di dati e dal tipo di applicazione a disposizione, mentre nel secondo ciclo (i) dalla relazione tra la quantità di dati e il miglioramento dell'algoritmo di targeting degli annunci, (ii) dalla relazione tra la qualità del targeting degli annunci e l'attrazione degli inserzionisti e (iii) dal modo in cui la piattaforma investe i ricavi pubblicitari e finanzia il miglioramento del servizio [23].

2.3.3 Esternalità di rete

In generale, in economia per esternalità si intende un qualsiasi effetto, positivo o negativo, che l'attività di produzione o di consumo di un soggetto ha sul benessere di un altro soggetto, senza che quest'ultimo paghi un prezzo (nel caso di impatto positivo) o riceva una compensazione (nel caso di impatto negativo).

Gli *effetti di rete* sono un tipo particolare di esternalità che ha luogo quando il valore dell'utilizzo di una tecnologia o di un servizio dipende dal numero di altre persone che lo usano. Per cui, l'utilità di un singolo agente i dipende non solo dalla quantità consumata q ma anche dal numero N di agenti nel mercato che lo usano, ovvero $U_i(q) = U_i(N, q)$.

Gli effetti di rete possono essere di due tipi:

Effetti di rete diretti: in cui il valore del servizio per gli utenti aumenta con il numero della stessa tipologia di utenti già attivi, ossia quando l'utilità di un agente di tipo A dipende dal numero di agenti appartenenti allo stesso gruppo A: $U_{i,A}(q) = U_{i,A}(N_A, q)$.

Ad esempio, il valore dell'adesione ad un social media o a un servizio di comunicazione è chiaramente in aumento nel numero di altri utenti con cui un individuo può potenzialmente interagire¹¹. Se si considerano, infatti, come gruppo A gli utilizzatori del servizio di messaggistica Whatsapp, la loro utilità dipende da quanti messaggi possono mandare tramite il servizio (q) e dal numero di persone che già lo usano (N_A).

In tal modo, un agente addizionale che entra nel network genera un'esternalità positiva (feedback positivo) per gli agenti esistenti poiché amplia le possibilità di "connessione". Questo ciclo di feedback positivo funziona anche al contrario: se l'adozione non raggiunge una massa critica di utenti, il bene o il servizio può cadere in una "spirale di morte" e alla fine scomparire.

Effetti di rete indiretti: in cui il valore che un utente deriva dall'adozione di un servizio è in aumento nel numero di utenti appartenenti ad un'altra categoria, ossia quando l'utilità di un agente appartenente ad un gruppo A dipende dal numero di agenti del gruppo B e viceversa: $U_{i,A}(q) = U_{i,A}(N_A(N_B), q)$ e $U_{i,B}(q) = U_{i,B}(N_B(N_A), q)$.

Questi tipi di effetti sono tipici delle piattaforme multi-versante, di cui si discuterà nel prossimo capitolo. In tal caso, infatti, il valore che gli utenti da un lato del mercato assegnano alla piattaforma dipende anche da quanti utenti su altri lati del mercato la patrocinano.

Ad esempio, i motori di ricerca beneficiano da un lato di effetti di rete indiretti positivi, poiché all'aumentare del numero di consumatori la piattaforma diventa più attraente per un altro gruppo, gli inserzionisti, dall'altro di effetti di rete indiretti negativi dagli inserzionisti ai consumatori in quanto questi ultimi in genere preferiscono i contenuti senza pubblicità.

In generale, gli effetti di rete sono spesso indicati come una fonte di concentrazione che facilita l'acquisizione del potere di mercato. Si sostiene che essi possano, infatti, portare a effetti *palla di neve* in cui i mercati si ribaltano a favore di uno o al massimo di un piccolo numero di player, cambiando in tal modo il tipo di competizione, che spesso è *per* il mercato (Schumpeterian) piuttosto che *nel* mercato [23]. In tali casi, dunque, i vincoli competitivi per gli operatori storici derivano dalla minaccia di essere "spostati" da un rivale o un entrante attraverso l'offerta di servizi radicalmente nuovi o migliorati, come quando una parte significativa dei consumatori è passata dal social network Myspace al concorrente innovativo e dirompente Facebook.

In particolare, per le piattaforme incumbent gli effetti di rete possono rappresentare un enorme vantaggio e potrebbero addirittura impedire a una piattaforma superiore di

¹¹ La legge di Metcalfe afferma che il valore di una rete è proporzionale al quadrato degli utilizzatori. Se in un certo istante ci sono n utilizzatori, l'ammontare complessivo delle possibili connessioni che si possono creare è $n(n-1) = n^2 - n$, approssimabile a n^2 per n molto grande.

sorpassarne una inferiore, per via della difficoltà per gli utenti di coordinare la migrazione a una nuova piattaforma. In effetti, anche se gli utenti stessero tutti meglio se migrassero *in massa* su una nuova e migliore piattaforma, non avrebbero necessariamente un incentivo *individuale* a farlo qualora avessero una bassa aspettativa che gli altri li seguirebbero. Si parla in tal caso di “incumbency advantage”, ossia di vantaggio dell’essere i first mover [40].

In conclusione, è spesso la combinazione di economie di scala, di scopo, di feedback loop e degli effetti di rete a far sì che gli operatori storici possano guadagnare e sfruttare il potere di mercato in un modo che non può essere eroso, almeno non sufficientemente prontamente, da potenziali concorrenti. Questo è il caso soprattutto in cui i consumatori preferiscono il single homing, cioè preferiscono, o per via di pregiudizi comportamentali o per via di costi di switching, quali i limiti di interoperabilità dei dati, instaurati dagli incumbent rivolgersi ad un solo provider per uno specifico servizio e non provare nuove alternative.

2.4 Fattori che attenuano il potere di mercato: il multi-homing

Per *multi-homing* si intende la pratica per la quale un utente utilizza più di un servizio dello stesso tipo attraverso fornitori diversi. Questa condizione può avere importanti implicazioni competitive. In particolare, il multi-homing indebolisce il ciclo di feedback, riduce l’importanza competitiva degli effetti di rete e di conseguenza diminuisce la propensione di un mercato a puntare su un operatore dominante, rendendo più probabile la coesistenza di più concorrenti.

Oggi per i principali servizi online, a differenza delle classiche industrie di rete, molto spesso gli utenti non devono sostenere dei costi per accedere ad altri fornitori di servizi. Ad esempio, per passare da un social network all’altro non è necessario acquistare un nuovo dispositivo o imparare ad operare in un sistema operativo diverso. Inoltre, i costi di utilizzo sono tipicamente bassi. In molti casi, infatti, questi servizi sono offerti gratuitamente.

Di conseguenza, da un punto di vista logico, il multi-homing sembrerebbe dover essere la norma più che l’eccezione nei mercati digitali. Gli utenti potrebbero idealmente patrocinare due piattaforme concorrenti allo stesso tempo. Nonostante questi presupposti, nella pratica esistono comunque dei costi di switching la cui entità varia da caso a caso. Se si considerano ad esempio i servizi di comunicazione, si ha che gli effetti di rete risultano semplicemente come un sottoprodotto dell’iscrizione/installazione del consumatore, per cui se questi installa sia WhatsApp che Telegram, può essere raggiunto attraverso entrambe le app. Al contrario, per i social network gli effetti di rete risultano come un sottoprodotto dell’utilizzo del servizio e non solo dall’iscrizione allo stesso. I social network, infatti, da un lato hanno bisogno che i loro utenti siiscrivano e dall’altro che curino i loro profili e si impegnino nel servizio. Questo è ciò che alcuni definiscono “multi-homing in uso”. Se l’utilizzo è costoso (ad esempio, perché curare e tenere aggiornato il proprio profilo richiede tempo), il semplice atto di multi-homing non cambia il fatto che in pratica i consumatori utilizzano ancora una

sola app. Dunque, dal punto di vista della concorrenza, si tratta di fatto di single-homing. Un caso illustrativo è Google+, il social network di Google. Nonostante Google abbia facilmente iscritto tutti i titolari di un account G Suite, la rete è stata alla fine interrotta nel 2019 per mancanza di impegno [41]. Questo esempio dimostra che anche in assenza di costi reali da sostenere per “switchare” ad un altro servizio, ci potrebbe essere ancora spazio per il vantaggio dell’incumbent grazie alla focalizzazione dei consumatori sul servizio offerto da quest’ultimo e alla poca disponibilità ad impegnarsi in servizi alternativi.

CAPITOLO 3

Big Data e Piattaforme Online

Come osservato nel paragrafo 1.4 del primo capitolo, nell'epicentro dell'ecosistema dei Big Data, le piattaforme online operano come principale interfaccia tra gli attori del mercato. In generale, esse rappresentano aziende che servono più gruppi di utenti contemporaneamente fornendo valore a ciascuno di essi in virtù della presenza dell'altro, e costituiscono oramai dei gateway fondamentali ai servizi innovativi che il mondo Internet mette a disposizione. In particolare, alcune di esse, per via della scala raggiunta, delle considerevoli valutazioni di mercato e delle enormi quantità di dati dei consumatori di cui dispongono, destano non poche preoccupazioni e rivestono nel panorama della questione dei Big Data un'influenza che non è solo economica ma si estende pure a questioni sociali e politiche, al punto tale da richiedere di esplorare come il diritto della concorrenza dovrebbe evolversi per garantire che l'era digitale continui a portare benefici ai consumatori.

Questo capitolo si occuperà di analizzare le loro caratteristiche, le loro strategie competitive e le implicazioni che queste determinano nell'analisi della concorrenza.

3.1 Definizione e tipi di piattaforme multi-versante

Le piattaforme multi-versante sono aziende che connettono due o più mercati, e dunque due o più gruppi di utenti, abilitando così la loro interazione e riducendo i costi di ricerca. Laddove, infatti, si individuano delle frizioni in un mercato che non rendono possibile una struttura efficiente dello stesso, le piattaforme intervengono come *matchmaker* al fine di superare le frizioni e generare valore per la società.

Più nello specifico, Evans e Schamlensee (2007) sostengono che una piattaforma multi-versante (che loro definiscono *catalizzatore economico*) “ha (a) due o più gruppi di clienti; (b) che hanno bisogno l'uno dell'altro in qualche modo; (c) ma che non possono catturare da soli il valore dalla loro reciproca attrazione; e (d) che si affidano al catalizzatore per facilitare le interazioni tra loro che creano valore” [43].

Il fulcro di questa definizione, che in sostanza cattura le caratteristiche chiave dell'attività della piattaforma, è il ruolo della stessa nella creazione di valore che non potrebbe esistere, o quantomeno sarebbe molto minore, in sua assenza. Questo valore è appunto generato dall'interazione di due (o più) gruppi distinti di utenti (versanti) garantita attraverso la piattaforma, la quale, in qualità di intermediario, coordina la loro domanda e ne favorisce la loro relazione tramite la riduzione dei costi di transazione¹².

¹² I costi di transazione sono i costi associati alla partecipazione ad un mercato e possono includere i costi di ricerca o informazione sostenuti per identificare opportunità rilevanti, il costo della negoziazione di accordi e il costo del trasferimento di un bene o servizio.

Esistono diversi tipi di piattaforme multi-versante [43]:

- **Piattaforme di scambio:** aiutano due gruppi di clienti, generalmente considerati “acquirenti” e “venditori”, a raggiungere un incontro in una situazione di reciproco vantaggio. Tra gli esempi vi sono le case d’asta, le agenzie immobiliari, i siti di e-commerce, le borse finanziarie, le agenzie di viaggio ecc.
- **Media supportati dalla pubblicità:** si tratta di piattaforme che creano o acquistano contenuti da altri allo scopo di attrarre gli spettatori, i quali a loro volta sono usati per attirare gli inserzionisti. È questo il caso di riviste, giornali, televisione gratuita, motori di ricerca, social network.
- **Sistemi di transazione:** si riferiscono a qualsiasi metodo di pagamento riconosciuto e utilizzato sia dagli acquirenti che dai venditori, come ad esempio le carte di credito.
- **Piattaforme software:** piattaforme che forniscono agli sviluppatori di applicazioni servizi quali l’accesso all’hardware per il dispositivo di comunicazione in questione e agli utenti che hanno la stessa piattaforma, la possibilità di eseguire queste applicazioni. Sono esempi i sistemi operativi per pc e smartphone, così come pure i videogiochi.

Il focus di questa discussione, sebbene il concetto di piattaforme multi-versante, come osservato negli esempi precedenti, non sia esclusivo dei mercati online, sarà su una classe specifica di piattaforme che fanno dell’uso dei Big Data uno degli elementi chiave dei loro modelli di business, ossia le **piattaforme online**, definite dalla Commissione Europea [43] come:

“imprese che operano in mercati a due (o più) lati, che utilizzano Internet per consentire interazioni tra due o più gruppi di utenti distinti ma interdipendenti in modo da generare valore per almeno uno dei gruppi”.

A tal proposito, tra gli esempi si rilevano piattaforme pubblicitarie online (AdSense, DoubleClick), mercati online (eBay, Amazon), motori di ricerca (Google, Bing), social media (Facebook, Instagram), piattaforme di distribuzione di applicazioni (Google Play e App Store), servizi di comunicazione (Facebook Messenger, Whatsapp), sistemi di pagamento (PayPal, Satyspay) e piattaforme per l’economia collaborativa (Uber, Airbnb).

Questa particolare categoria di piattaforme, che ha acquisito un’importanza crescente nell’economia digitale, dispone di alcune comuni e importanti caratteristiche specifiche. In particolare, secondo la Commissione Europea [44], esse:

- Hanno la capacità di creare e plasmare nuovi mercati, di sfidare quelli tradizionali e di organizzare nuove forme di partecipazione o di condurre attività economiche basate sulla raccolta, elaborazione e uso di grandi quantità di dati;

- operano in mercati a più lati, ma con diversi gradi di controllo sulle interazioni dirette tra gruppi di utenti;
- beneficiano di "effetti di rete", per cui, in generale, il valore del servizio aumenta con il numero di utenti;
- spesso si affidano alle tecnologie dell'informazione e della comunicazione per raggiungere i loro utenti, facilmente e istantaneamente;
- svolgono un ruolo chiave nella creazione di valore digitale, in particolare catturando un valore significativo (anche attraverso l'accumulo di dati), facilitando nuove iniziative imprenditoriali e creando nuove dipendenze strategiche.

3.2 Caratteristiche delle piattaforme multi-versante

3.2.1 Effetti di rete

Una qualsiasi piattaforma multi-versante, poiché permette l'interazione di almeno due tipi distinti di utenti, è caratterizzata da quattro effetti di rete, come si osserva in figura 3.1:

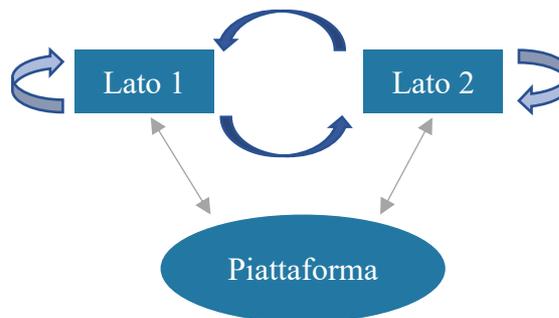


Figura 3.1: Effetti di rete diretti e indiretti in un mercato a due lati

- Un effetto di rete diretto (o *same-side effect*) per ogni lato, per cui il valore della piattaforma varia al variare del numero di utenti sullo *stesso* lato.
- Un effetto di rete indiretto (o *cross-side effect*) in ciascuna direzione, per cui il valore della piattaforma varia al variare del numero di utenti sull'*altro* lato.

Ognuno di questi effetti può essere positivo, qualora un utente aggiuntivo aumenti il valore della piattaforma, o negativo se lo diminuisce.

Ad esempio, su un social network, classico caso di una piattaforma supportata dalla pubblicità, se da un lato gli inserzionisti accolgono con favore la partecipazione di nuovi consumatori sulla piattaforma (poiché ciò aumenta la possibilità di comunicare con un potenziale cliente, la prospettiva di maggiori visualizzazioni e quindi di vendite), dall'altro non gradiscono la crescita nei loro mercati, ossia la partecipazione di altri inserzionisti concorrenti. Per quanto riguarda i consumatori, invece, questi apprezzano sicuramente il fatto di avere più persone con cui comunicare, ma al contrario, solitamente, non accolgono con favore un numero eccessivo di pubblicità e quindi di inserzionisti. Così, nel presente esempio, gli inserzionisti beneficiano di un effetto di rete cross-side positivo mentre

dispongono di un effetto sullo stesso lato probabilmente negativo. Il contrario vale per i consumatori.

La presenza di queste esternalità di rete indirette ha un'importante implicazione per l'analisi economica delle piattaforme multi-versante. Infatti, se per le industrie tradizionali a un solo lato gli economisti partono normalmente dal presupposto che la domanda dipenda dal prezzo del prodotto e dai prezzi dei complementi e dei sostituti, per le piattaforme a più lati la domanda da parte di un gruppo di utenti dipende anche dalla dimensione di ciascuno degli altri gruppi di agenti economici che la piattaforma serve. In particolare, in presenza di effetti di rete positivi, l'interdipendenza tra i diversi gruppi di utenti è tale che “un maggiore coinvolgimento di utenti di almeno un tipo aumenta il valore della piattaforma per gli agenti di altri tipi” [42]. In tale scenario, la piattaforma ha un ruolo chiave nel riconoscere l'esistenza di queste interdipendenze e quindi nell'internalizzarle. In tale modo, la piattaforma gode di un vantaggio rispetto alle imprese che servono l'uno o l'altro mercato separatamente, poiché internalizzando l'impatto sociale che l'uso di un prodotto o di un servizio di un gruppo ha sull'altro gruppo, è capace di fissare i prezzi in modo più efficace in modo da far crescere i mercati abbinati.

3.2.2 Pricing asimmetrico

Gli effetti di rete incrociati forniscono alle piattaforme un incentivo ad adottare strutture di prezzo insolite che comportano delle *sovvenzioni* o *sussidiazioni incrociate*.

Più nel dettaglio, le piattaforme, consapevoli del valore che l'interazione dei due gruppi genera, si pongono come obiettivo quello di “portare a bordo” entrambi i lati. Di conseguenza, poiché i prezzi applicati condizionano la presenza e l'adesione degli utenti sulla piattaforma e impattano sulla volontà delle due parti di interagire una volta su di essa, è fondamentale che la struttura dei prezzi sia opportunamente definita a tal scopo.

Gli stessi Rochet e Tirole (2006) definiscono un mercato bifronte come “un mercato in cui il volume delle transazioni tra gli utenti finali dipende dalla *struttura* e non solo dal livello complessivo *delle tariffe* applicate dalla piattaforma” [45]. Si parla in tal caso di “non neutralità” della struttura dei prezzi: il modo in cui i prezzi sono distribuiti tra gli utenti dei due versanti influisce quindi sulla domanda e sul livello di transazione.

In particolare, ciò comporta la tendenza delle piattaforme a stabilire una *struttura dei prezzi asimmetrica*, tale per cui i ricavi ottenuti da un tipo di utenti sovvenzionano in sostanza l'altro gruppo (versante sussidiato), al quale viene addebitato un prezzo relativamente basso, potenzialmente inferiore al costo marginale, o addirittura un prezzo nullo.

Le particolari caratteristiche delle piattaforme sono quindi tali per cui questa scelta non costituisce necessariamente una fattispecie abusiva, ma al contrario potrebbe rientrare - anche nel caso di piattaforma monopolista - nella logica ottimale di massimizzazione del profitto delle imprese. In tal caso, quindi, non vi sarebbe alcun motivo, oltre che dal punto di vista dei profitti della piattaforma anche dal punto di vista del benessere sociale, per cui i prezzi applicati alle due parti riflettano i rispettivi costi di fornitura dei servizi.

Come si osserva nella tabella 3.1 di seguito, tratta da Parker e Van Alstyne (2005) [46], per ogni mercato bilaterale servito da una piattaforma in qualità di intermediario esiste sempre un lato sussidiato dall'altro. Il lato sussidiato è trattato come un leader di perdita a cui si applica quindi un prezzo inferiore al costo marginale, l'altro lato, al contrario, è trattato come centro di profitto a cui si applica un prezzo invece superiore al costo marginale.

Tabella 3.1: Esempi di mercati two-sided [46]

Product category	Market 1	Intermediary	Market 2
Portable documents	Document reader [Ⓜ]	Adobe	Document writer
Credit cards	Consumer credit [Ⓜ]	Issuing bank	Merchant processing
Operating systems	Complementary applications	Microsoft, Apple, Sun	Systems developer toolkits [Ⓜ]
Plug-ins	Applications software	Microsoft, Adobe	Systems developer toolkits [Ⓜ]
Ladies' nights	Men's admission	Bars, restaurants	Women's admission [Ⓜ]
TV format	Color UHF, VHF, HDTV [Ⓜ]	Sony, Phillips, RCA	Broadcast equipment
Broadcast & publishing	Content [Ⓜ]	Magazine publishers, TV, radio broadcasters	Advertisements
Computer games	Game engine/player	Ubisoft, ID, valve, electronic arts	Level editors [Ⓜ]
Auctions	Buyers [Ⓜ]	E-Bay, Christie's, Sotheby's	Sellers
Academic journals	Articles	<i>Management Science</i>	Author submissions [Ⓜ]
Recruiting	Applicants [Ⓜ]	Monster.com	Employers
Reservation systems	Travelers [Ⓜ]	Expedia, Travelocity, Orbitz	Hotels, airlines, rental cars
Shopping malls	Shoppers [Ⓜ]	Mall of America	Stores
Streaming audio/video	Content [Ⓜ]	Real audio, Microsoft, Apple	Servers
Paid search	Searchers [Ⓜ]	Google.com	Marketers
Stock exchange	Equity purchasers [Ⓜ]	NYSE, NASDAQ	Listed companies
Home real estate	Home buyers [Ⓜ]	Real estate agents	Home sellers

Notes. This table shows how one side of a two-sided network market receives a discounted, free, or even subsidized good (indicated with [Ⓜ]). In general though not always, Market 1 can be interpreted as the user/consumer market and Market 2 can be interpreted as the producer/developer market. We provide a test for which side receives the free good below.

Ad esempio, il modello di business della TV non a pagamento, così come quello dei servizi web, quali i motori di ricerca, i social network, l'e-mail, è quello di trattare i consumatori come leader di perdita, usarli per attirare gli inserzionisti e trattare dunque questi ultimi come centro di profitto. Al contrario nell'industria dei PC per i sistemi operativi, sono gli utenti finali, che acquistano una piattaforma che consiste in un sistema operativo fornito in bundle con l'hardware, ad essere considerati come centro di profitto e gli sviluppatori di applicazioni, invece, come leader di perdita, ovvero sono sul lato dei sussidi in quanto non pagano royalty e ricevono kit di sviluppo software gratuiti dai fornitori di sistemi operativi.

Questa strategia ha senso per via degli effetti di rete incrociati positivi che possono sussistere. Infatti, per un'attività ad un solo lato, applicare un prezzo nullo o comunque inferiore al costo marginale significherebbe rinunciare a tutti i profitti su ogni unità venduta. La situazione è diversa per le piattaforme che servono mercati multi-versante caratterizzati da effetti di rete incrociati tali per cui la perdita su un lato del mercato può essere più che compensata dai profitti che si fanno sull'altro lato.

Si consideri, ad esempio, il mercato Acrobat PDF di Adobe che comprende sia i lettori che gli scrittori di documenti. Supponendo che l'azienda applichi un prezzo positivo e superiore

ai costi marginali ad entrambi i lati, i profitti, come si osserva in figura 3.2, saranno realizzati in entrambi i mercati.

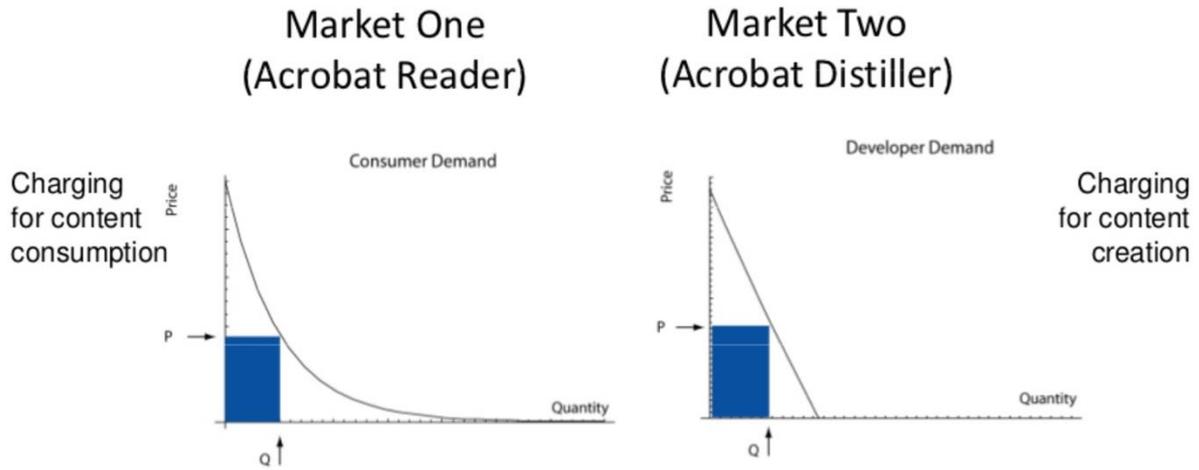


Figura 3.2: Curve di domande e profitti con pricing tradizionale

Tuttavia, è noto che l'azienda applica un altro modello di business: i lettori non pagano nulla per il software Acrobat, al contrario dei produttori di documenti che pagano per l'accesso ad un'ampia base di lettori. Sovvenzionando il mercato uno, come si evince dalla figura 3.3, si produce un aumento delle adesioni e dei consumi da quel lato, che per via dell'effetto rete stimola la crescita della domanda nel mercato adiacente. In tal modo, dunque, l'aumento dei profitti nel mercato due sarà più che sufficiente a coprire le perdite nel mercato uno.

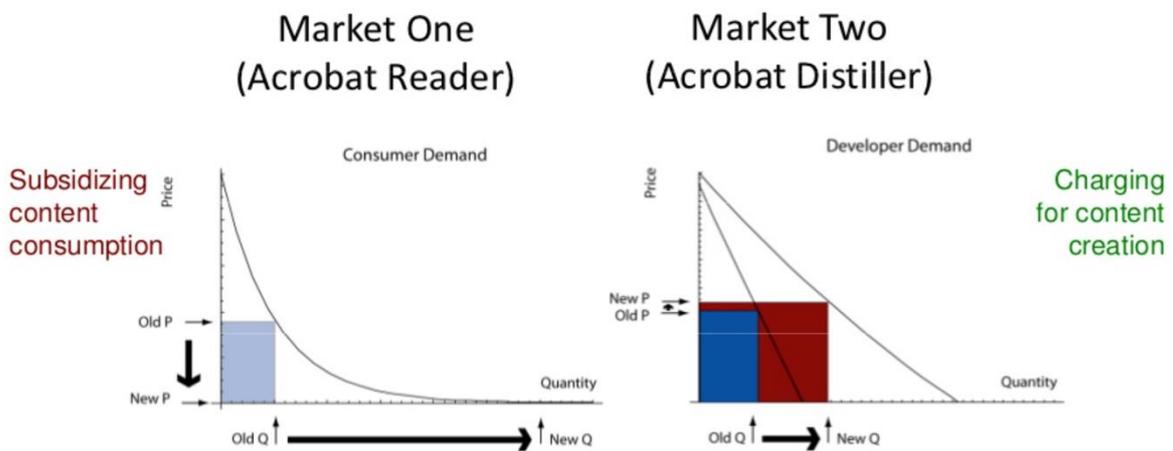


Figura 3.3: Curve di domande e profitti con pricing asimmetrico

In conclusione, l'effetto rete incrociata fornisce ad Adobe un incentivo a sovvenzionare in questo modo. Facendo pagare ai lettori un prezzo nullo, l'azienda di software assicura che un numero elevato di lettori utilizzi il suo servizio, aumenta il valore del programma per i produttori di documenti, il che a sua volta aumenta i ricavi che la piattaforma può guadagnare da quest'altro gruppo.

Nella determinazione della struttura dei prezzi tale da coinvolgere entrambe le parti, la scelta del giusto lato da subsidiare è cruciale per sviluppare forti effetti di rete. La sfida per una piattaforma è, dunque, quella di determinare in *che misura* e *sulla base di quali fattori* un gruppo dovrebbe essere incoraggiato alla partecipazione attraverso il sovvenzionamento.

In due modelli diversi, Parker e Van Alstyne (2005) e Rochet Tirole (2006), si dimostra che la scelta del lato del mercato two-sided da subsidiare dipende dalle dimensioni relative degli effetti di rete e dall'elasticità dei prezzi delle due domande.

3.2.2.1 Modello di Parker e Van Alstyne (2005)

Parker e Van Alstyne considerano un generico mercato a due lati con competizione sul prezzo. Le domande dirette dei due lati sono date da:

$$D(p_1) = 1 - p_1 \quad D(p_2) = 1 - p_2$$

Tuttavia, in presenza di esternalità di rete indirette, la domanda di un lato dipende, oltre che dal prezzo ad esso applicato, anche dal prezzo applicato all'altro versante, per cui:

$$q_1 = D(p_1) + e_{21}D(p_2) = 1 + e_{21}D(p_2) - p_1$$

$$q_2 = D(p_2) + e_{12}D(p_1) = 1 + e_{12}D(p_1) - p_2$$

Con $e_{ij} \geq 0$ e $e_{12}e_{21} < 1$.

Dove e_{ij} rappresenta un effetto di attrazione, esternalità di rete, che il gruppo i genera nei confronti del gruppo j . Più precisamente, e_{ij} misura l'effetto che una variazione della domanda nel mercato i ha sul mercato j .

In tal modo se $D(p_2)$ aumenta (per via di una diminuzione di p_2) si ha un effetto positivo anche sulla domanda del mercato 1, che cresce di $e_{21}D(p_2)$.

Si suppone che il mercato multi-versante sia servito da una *piattaforma monopolista* che dunque internalizza le esternalità su ambo i lati. La piattaforma determinerà i prezzi da applicare ai due gruppi sulla base della massimizzazione del profitto congiunto, il quale, lasciando che i costi marginali per i beni informativi siano trascurabili con il profitto, è dato da:

$$\text{Max}_{p_1, p_2} \pi = \pi_1 + \pi_2 = p_1 q_1 + p_2 q_2$$

Risolvendo le due FOC $\frac{\delta \pi}{\delta p_1} = 0$ e $\frac{\delta \pi}{\delta p_2} = 0$, si ottengono le curve di reazione seguenti:

$$\begin{cases} p_1(p_2) = \frac{1 + e_{21} - p_2(e_{21} + e_{12})}{2} \\ p_2(p_1) = \frac{1 + e_{12} - p_1(e_{12} + e_{21})}{2} \end{cases}$$

Da cui si evince che più il prezzo di un lato aumenta più quello dell'altro lato diminuisce e viceversa. All'equilibrio si avrà:

$$p_1^* = \frac{1 - e_{12}}{2 - (e_{12} + e_{21})} ; \quad p_2^* = \frac{1 - e_{21}}{2 - (e_{12} + e_{21})} ; \quad \pi^* = \frac{1 - e_{12}e_{21}}{2 - (e_{12} + e_{21})}$$

Dai prezzi di equilibrio si deduce che il prezzo applicato a ciascun lato dipende inversamente dall'esternalità che questi genera sull'altro gruppo.

In particolare, si possono distinguere, come si osserva in figura 3.4, tre casi caratterizzati da diversi livelli di esternalità indiretta:

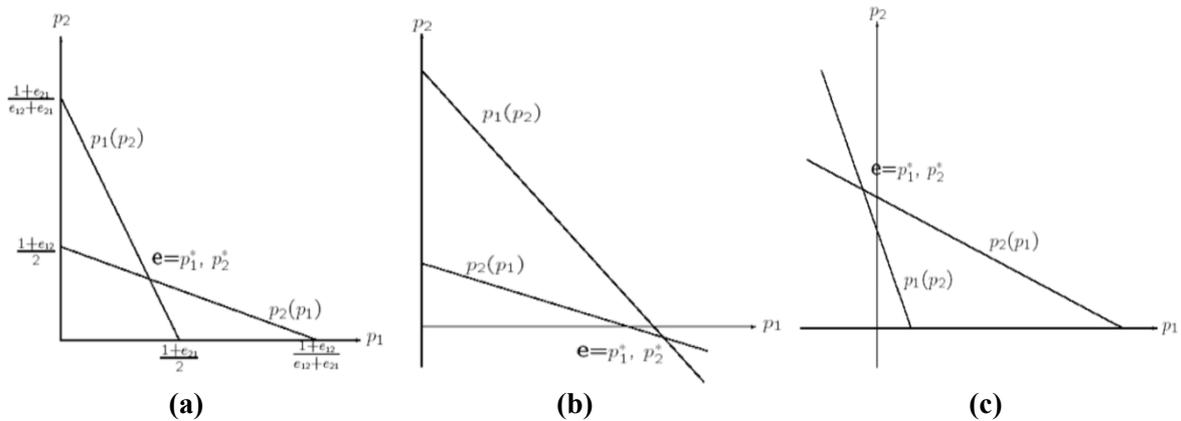


Figura 3.4: Curve di reazione in presenza di livelli diversi di esternalità indiretta

I prezzi di equilibrio corrispondono al punto di intersezione delle due curve.

- (a) Con un **livello intermedio di esternalità indiretta** si ha che ciascun lato della piattaforma genera un considerevole beneficio per l'altro lato, per cui entrambi pagheranno un prezzo positivo.
- (b) In presenza di **alta esternalità da 2 a 1 ($e_{21} > e_{12}$)** si avrà che tanti più utenti del gruppo 2 aderiscono alla piattaforma tanto maggiore sarà l'attrazione di utenti nel gruppo 1. Il gruppo 2 è quello che dà un beneficio molto più alto di quello che dà il gruppo 1, per cui sarà ottimale dal punto di vista della massimizzazione del profitto, fissare un prezzo positivo per il gruppo 1 e negativo per il gruppo 2.
- (c) In presenza di **alta esternalità da 1 a 2 ($e_{21} < e_{12}$)** si avrà, invece, che tanti più utenti del gruppo 1 aderiscono alla piattaforma tanto maggiore sarà l'attrazione di utenti del gruppo 2. Il gruppo 1 è quello che dà un beneficio molto più alto di quello che dà il gruppo 2, per cui sarà ottimale dal punto di vista della massimizzazione del profitto, fissare un prezzo positivo per il gruppo 2 e negativo per il gruppo 1.

In definitiva, si dimostra che nei mercati delle piattaforme le strategie di prezzo sono dunque fortemente condizionate in funzione di quale gruppo genera più attrazione: il versante che pagherà di meno attraverso un sussidio sarà quello che genera un effetto di rete maggiore sull'altro lato.

3.2.2.2 Modello di Rochet e Tirole (2006)

Rochet e Tirole (2006) prendono in considerazione il mercato delle carte di credito che vede da un lato i titolari delle carte di credito disposti a pagare con esse, e dall'altro i commercianti disposti ad accettarle come metodo di pagamento.

Si suppone che tale mercato bi-laterale sia servito da una piattaforma monopolista che sostiene un costo fisso nullo per servire ogni membro su un qualsiasi lato ed un costo marginale $c \geq 0$ per l'interazione tra due membri di lati opposti.

Per ciascun lato, i membri pagano una tariffa d'uso per transazione rappresentata da p^m e p^c , rispettivamente per i commercianti (*merchants*) e per i consumatori (*consumers*).

La richiesta di transazioni da parte di un gruppo è data da $D^i(p^i)$, con $i=m,c$, ed è dipendente appunto dalla tariffa d'uso applicata dalla piattaforma. Inoltre, la domanda è rappresentata dal numero di transazioni potenziali date dal prodotto delle richieste dei singoli gruppi, ossia $D^m(p^m) \cdot D^c(p^c)$.

In queste condizioni, la piattaforma monopolista, determinerà i prezzi da applicare a ciascun lato massimizzando il profitto congiunto:

$$\text{Max}_{p^c, p^m} \quad \pi = (p^c + p^m - c) D^m(p^m) D^c(p^c)$$

Risolvendo le FOC $\frac{\delta \pi}{\delta p^c} = 0$ e $\frac{\delta \pi}{\delta p^m} = 0$ si trova:

$$p^c + p^m - c = \frac{p^c}{\varepsilon^c} = \frac{p^m}{\varepsilon^m}$$

Con ε^i che rappresenta l'elasticità della domanda del lato i .

Questa formula può essere riscritta come una formula di Lerner standard:

$$\frac{(p^c - (c - p^m))}{p^c} = \frac{1}{\varepsilon^c} \quad \frac{(p^m - (c - p^c))}{p^m} = \frac{1}{\varepsilon^m}$$

Dove $c - p^m$ è il costo opportunità per il gruppo c e $c - p^c$ è il costo opportunità per il gruppo m , e rappresentano una sorta di costo residuo da coprire una volta incassato il prezzo dall'altro lato.

Infatti, quando non ci sono costi e benefici fissi, la perdita di una transazione sul lato i dovuta ad un aumento del prezzo per transazione p^i ha un costo opportunità $c - p^j$ poiché

il costo per transazione c della piattaforma deve essere compensato dal pagamento p^j prelevato dall'altro lato.

Di conseguenza, il prezzo che la piattaforma fissa per un lato del mercato sarà tanto più basso tanto maggiore sarà il prezzo applicato all'altra parte perché minore saranno i costi da coprire. Inoltre, il lato sussidiato sarà quello caratterizzato da una domanda più elastica.

Dopo aver considerato i precedenti due modelli si può affermare che, allo scopo di coinvolgere entrambi i lati di un mercato two-sided e di massimizzare il profitto, le piattaforme progettano la loro struttura dei prezzi sulla base dei due seguenti fattori:

- *Gli effetti di rete incrociati e le loro dimensioni relative*: ha senso sovvenzionare il lato che genera un effetto di attrazione maggiore, e far pagare di più il lato che aumenta la sua domanda in modo più forte in risposta alla crescita della domanda dell'altro lato.
- *Elasticità al prezzo dell'utente*: ha senso sovvenzionare il lato più sensibile al prezzo perché altrimenti gli utenti appartenenti ad esso non sarebbero incentivati ad aderire alla piattaforma. Ad esempio, se Adobe invertisse il suo approccio, facendo pagare ai lettori e sovvenzionando gli scrittori, la sua rete crollerebbe. I produttori di documenti, infatti, sono meno sensibili al prezzo, quindi il software libero non aumenterebbe drammaticamente il loro numero e in più apprezzano molto un enorme pubblico. Al contrario, dall'altro lato, i lettori sono molto più sensibili al prezzo e non pagherebbero molto per l'accesso a una base più ampia di scrittori [47].

Inoltre, occorre dire che le decisioni sui prezzi sono più semplici quando ogni nuovo utente sovvenzionato non costa praticamente nulla al fornitore della piattaforma, cosa che accade quando l'offerta assume la forma di un bene digitale. Tuttavia, quando un prodotto offerto ha costi unitari apprezzabili, come nel caso dei beni materiali, i fornitori della piattaforma devono essere più attenti poiché in questo caso se sul lato "pagante" non si materializza una forte disponibilità a pagare, una strategia di sussidiazione con alti costi variabili può rapidamente accumulare grandi perdite [47].

Proprio la proprietà a costo marginale quasi nullo dell'informazione, tipica dei servizi online, permette molto spesso ad una piattaforma digitale di sovvenzionare un mercato arbitrariamente grande ad un costo modesto.

3.3 Impatto economico delle piattaforme online

Nell'ultimo decennio le piattaforme online sono cresciute notevolmente in scala e dimensioni, diventando sempre di più aziende dirompenti rispetto agli approcci tradizionali e introducendosi in altri settori, che vanno dai trasporti alle banche, oltre a quello dei social media, dei viaggi, dei libri.

Esse costituiscono un contributo importante per le economie, innanzitutto perché, poiché rendono possibile in alcuni casi un matching altamente efficiente (si pensi ai mercati dell'e-

commerce come eBay o alle reti professionali come LinkedIn) ed in altri forniscono un supporto ad un utilizzo più efficiente delle risorse (come nel caso di Airbnb, Uber e in altri esempi della “share economy”), rappresentano una fonte di aumento della produttività.

In secondo luogo, le piattaforme sono state importanti fonti di innovazione, basti pensare che nel 2014 nove piattaforme statunitensi hanno ottenuto 11.585 brevetti. Infine, molte piattaforme start-up sono riuscite ad attrarre investimenti significativi da parte di fondi di venture capital, tant'è vero che la maggior parte dei cosiddetti "unicorni"¹³ sono infatti società di piattaforme [48].

Inoltre, i modelli di business delle piattaforme online si sono dimostrati altamente redditizi al punto tale da permettere ad alcune di esse di essere tra le prime dieci aziende più grandi al mondo per capitalizzazione di mercato¹⁴.

Uno dei tentativi più rilevanti di descrivere l'ascesa globale delle piattaforme online è l'indagine di Evans e Gawer (2016) [48] che si concentra su piattaforme sia private che quotate in borsa con una valutazione di mercato di almeno un miliardo di dollari. Tra le 176 aziende così selezionate è risultato che: 82 erano in Asia (prevalentemente in Cina), 64 in Nord America (tutte con sede negli Stati Uniti ad eccezione di una canadese), 27 in Europa (un terzo nel Regno Unito), 2 in America Latina (Argentina e Brasile) e 1 in Africa (Sud Africa).

La loro capitalizzazione di mercato globale superava i **4,3 trilioni di dollari**, a dimostrazione del valore che le aziende di piattaforma hanno raggiunto negli ultimi anni. Sebbene sia l'Asia ad avere il maggior numero di piattaforme, è il valore delle piattaforme provenienti dal Nord America, a cui corrisponde una capitalizzazione complessiva di 3,1 trilioni di dollari, a fornire il maggior contributo al valore complessivo (72% contro il 22% per quelle asiatiche). Altro elemento da sottolineare è che sorprendentemente, mentre l'Europa è emersa come uno dei principali consumatori di servizi di piattaforma, ne ha generate relativamente poche (solo il 15%) che nel complesso rappresentano poco più del 4% del valore di mercato.

Sulla base del modello di business adottato, le piattaforme online sono state classificate in quattro tipologie principali [48]:

- Le **piattaforme di transazione** che fanno da intermediario tra diversi utenti, acquirenti e fornitori, velocizzando e facilitando lo scambio (ad esempio Uber e Airbnb). Ne fanno parte 160 piattaforme di tale indagine su 176.

¹³ Gli Unicorns sono start-up private che hanno raggiunto una valutazione di 1 miliardo di dollari o più senza quotarsi in borsa. Da un'analisi delle 115 società elencate come Unicorns da CB Insights nel giugno 2015 è emerso che 80 di queste società, ovvero il 70%, sono società di piattaforma [48].

¹⁴ Aziende quali Microsoft, Apple, Amazon, Alphabet, Facebook e Alibaba ricoprono rispettivamente il secondo, terzo, quarto, quinto, sesto, settimo posto nella classifica delle più grandi società del mondo per capitalizzazione di mercato nel 2020. <https://www.statista.com>

- Le **piattaforme di innovazione** le quali sono aziende che sviluppano un prodotto o un servizio su cui altre imprese costruiscono tecnologie, prodotti o servizi complementari (ad esempio Microsoft e Oracle). Ne fanno parte 5/176 piattaforme.
- Le **piattaforme integrate**, che integrano le caratteristiche delle piattaforme precedenti (ad esempio Google e Facebook); Ne fanno parte 6/176 piattaforme.
- Le **piattaforme di investimento** sono aziende che hanno sviluppato una strategia di portafoglio della piattaforma e agiscono come una holding (ad esempio SoftBank e Naspers). Ne fanno parte 5/176 piattaforme.

Come si osserva in figura 3.5, sebbene oltre il 90% delle piattaforme incluse nell'indagine sono da ricondursi alla prima categoria, sono le piattaforme integrate a contribuire maggiormente al valore di mercato globale. Infatti, queste ultime, in appena sei, raggiungono un tetto di mercato totale di 2.000 miliardi di dollari.

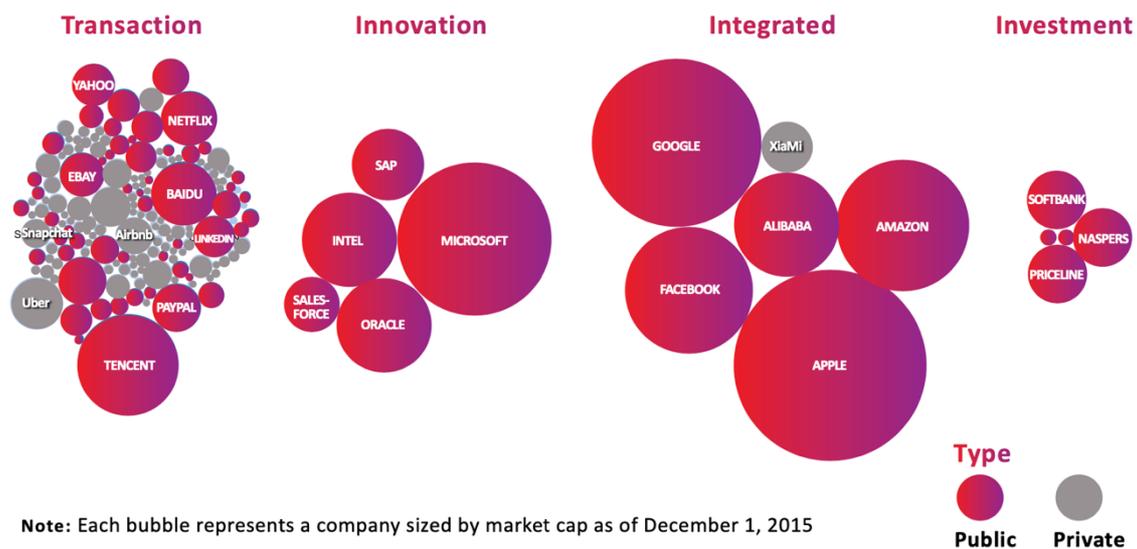


Figura 3.5: Società di piattaforme classificate in base al tipo [48]

Dall'indagine è inoltre emerso che, considerando le prime 25 principali piattaforme online, la somma delle valutazioni di mercato delle prime cinque società (Apple, Google, Microsoft, Amazon e Facebook) supera visibilmente la valutazione di tutte le altre venti società nel loro insieme.

3.4 Le piattaforme di attenzione

Un tipo particolare di piattaforme online è rappresentato dalle *piattaforme (o broker) di attenzione*, le quali sono aziende capaci di (i) ottenere informazioni sulle preferenze individuali degli utenti e (ii) indirizzare singolarmente, sulla base di queste, gli annunci pubblicitari [49].

Piattaforme quali i motori di ricerca e i social network ne sono un esempio. Esse, infatti, organizzano servizi e contenuti nel tentativo di facilitare il processo di ricerca dei consumatori finali allo scopo di aggregare la loro attenzione. Una volta che la piattaforma avrà aggregato abbastanza attenzione, cioè quando sarà considerata "utile" da un gran numero di consumatori, allora potrà monetizzare il suo ruolo di "information gatekeeper" vendendo agli inserzionisti l'accesso all'attenzione dei consumatori, ovvero la possibilità di inserire pubblicità mentre questi utilizzano la piattaforma [50].

Poiché l'attenzione dei consumatori è alla base di questo modello di business ed è, inoltre, una risorsa scarsa, nel complesso gli interessi delle piattaforme di attenzione sono allineati con la massimizzazione, *almeno nel breve termine*, del surplus dei loro consumatori [50]. La ragione fondamentale è che, per via della sua scarsità, l'attenzione dei consumatori è molto elastica. Se una data piattaforma, infatti, non è percepita come sufficientemente utile (ad esempio, perché le informazioni ottenute attraverso essa sono obsolete o imprecise, i prodotti e i servizi intermediati sono troppo costosi, o la varietà di prodotti, servizi o informazioni che possono essere ottenuti attraverso la piattaforma è troppo limitata), allora i consumatori ri-ottimizzeranno la loro allocazione dell'attenzione, riducendo l'attenzione data a questa piattaforma, e aumentando l'attenzione data ad altre (non necessariamente nello stesso dominio), o forse anche ad altre attività (offline). In breve: *l'elasticità dell'attenzione dei consumatori limita fortemente il canone che le piattaforme possono estrarre dai consumatori*. All'estremo, una piattaforma potrebbe essere sostituita da una nuova piattaforma che soddisfa lo stesso scopo, ma lascia ai consumatori un surplus di consumo più elevato [50].

Di conseguenza le piattaforme di attenzione concorrono, dunque, per i bulbi oculari offrendo nuove funzionalità, servizi preziosi, innovativi e spesso gratuiti per attirare tanti più spettatori e mantenerli sulla piattaforma. In linea con la logica del pricing asimmetrico delle piattaforme multi-versante definita precedentemente, questi servizi gratuiti sono sovvenzionati innanzitutto tramite i profitti estratti all'altro lato del mercato, cioè agli inserzionisti, la cui domanda, appunto perché devono seguire le (poche) piattaforme che aggregano l'attenzione dei consumatori, è molto meno flessibile [50].

3.4.1 Ruolo dei Big Data: il costo del “gratuito”

“Iscriviti. È gratis e lo sarà sempre”: questo era il messaggio che, fino a gennaio 2019, un utente trovava sul form di registrazione al social network Facebook.

Non è insolito, infatti, che le piattaforme di attenzione promuovano e offrano ai consumatori servizi “gratuiti”. Tale strategia, sebbene come osservato precedentemente, rientri nella logica ottimale di massimizzazione del profitto per una piattaforma multi-versante, potrebbe indurre in errore i consumatori e creare in loro l'impressione che l'utilizzo del servizio offerto sia assolutamente privo di costi.

In realtà, invece, il fatto che le piattaforme di attenzione non chiedano ai consumatori un prezzo monetario per accedere ai loro servizi non esclude che essi paghino comunque un costo non pecuniario per farlo. Oltre alla loro attenzione, infatti, i consumatori finali, in

cambio dei numerosi e variegati servizi offerti dalle piattaforme, trasferiscono (e di fatto "pagano" con)¹⁵ i propri dati (cfr. figura 3.6). Per tale motivo, il termine "gratuito" in questi casi è oltre che fuorviante, inappropriato.



Figura 3.6: Scambi di valore nella fornitura di servizi a costo monetario zero [51].

È ovvio che per un'azienda che dipende dagli utenti che rimangono online per guardare gli annunci è fondamentale offrire loro contenuti mirati. Per farlo deve alimentare i propri algoritmi con i dati degli utenti affinché questi imparino su quali contenuti le persone cliccheranno, e quali contenuti le faranno rimanere sulla piattaforma più a lungo. Questi dati, come già osservato, costituiscono quindi un input fondamentale per rendere più efficace la pubblicità attraverso la profilazione.

I consumatori in sostanza barattano la loro privacy e le informazioni circa ad esempio i beni che vorrebbero acquistare, i luoghi che vorrebbero visitare, i ristoranti in cui vorrebbero mangiare, in cambio di servizi digitali "gratuiti". In tal modo sicuramente l'utente ha il vantaggio immediato del servizio a prezzo zero, ma molto spesso non è sufficientemente consapevole dei costi a breve o a lungo termine che la divulgazione delle sue informazioni comporterà, in quanto potrebbe non conoscere nel dettaglio come i dati saranno utilizzati e da chi [51]. In definitiva, i dati degli utenti possono anche essere considerati come beni per le piattaforme digitali che possono essere venduti, concessi in licenza, divulgati o scambiati con terzi.

3.4.2 Oligopolio dell'attenzione

Tra tutte le aziende che rientrano nella definizione di piattaforme di attenzione data precedentemente (ovvero piattaforme capaci di (i) ottenere informazioni sulle preferenze dei singoli utenti e (ii) indirizzare singolarmente, sulla base di queste, gli annunci pubblicitari), sicuramente Google e Facebook sono le più preziose. Google costituisce il mezzo principale con cui le persone nel mondo occidentale trovano informazioni e contenuti su Internet, mentre Facebook, con oltre 2,5 miliardi di utenti, è il mezzo principale con cui le persone si connettono e comunicano tra loro.

Grazie alla loro migliore capacità di apprendere le preferenze dei consumatori, attraverso una capacità di raccolta e analisi dei Big Data incomparabile per altre aziende, riescono ad aggregare molta attenzione dei consumatori e a distribuire loro annunci iper-mirati.

¹⁵ Alcuni commentatori sostengono che i consumatori "pagano con i loro dati"; si veda, ad esempio, il discorso del commissario CE per la concorrenza Vestager https://ec.europa.eu/commission/commissioners/2014-2019/vestager/announcements/making-data-work-us_en.

Insieme, infatti, costituiscono di fatto un vero e proprio *duopolio dell'attenzione*. A tal proposito, basta considerare che la maggior parte della spesa pubblicitaria digitale in tutto il mondo (70%), come si osserva in figura 3.7, va a Google (46%) e a Facebook (24%). Anche se si considerano le entrate pubblicitarie su tutti i media, il duopolio digitale cattura oltre un terzo di tutti i dollari spesi nel 2019.

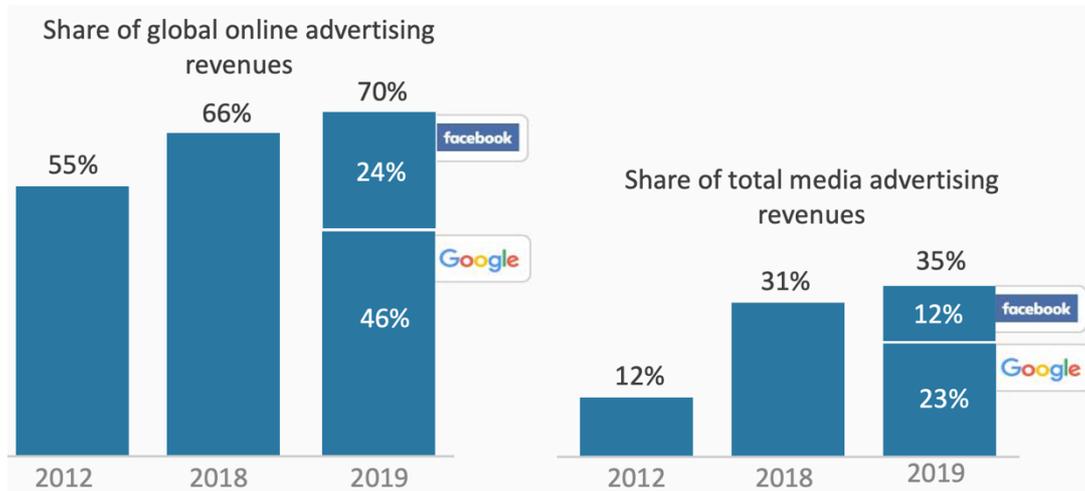


Figura 3.7: Quote di mercato di Google e Facebook rispetto ai ricavi globali della pubblicità online e della pubblicità sui media in generale. Fonte: www.statista.com

Questa loro condizione gli consente di avere un alto grado di controllo e di influenza sull'accesso da parte degli inserzionisti ai potenziali acquirenti e per tale motivo sono considerate come un fondamentale, quasi obbligatorio, gatekeeper tra le imprese e i loro potenziali clienti. Di conseguenza le due piattaforme sollevano non poche preoccupazioni riguardo a come il loro “*potere del collo di bottiglia*” potrebbe influire sui mercati in cui operano.

Infatti, come osservato nella sezione 3.4, poiché l'attenzione continua dei consumatori è alla base dei modelli di business di questa particolare categoria di piattaforme, queste faranno di tutto per offrire ai consumatori servizi innovativi e gratuiti, per cui sicuramente il loro interesse è allineato con il surplus del consumatore *a breve termine*. Nonostante questo, però, ci sono due questioni che possono destare preoccupazioni: in primo luogo come il *surplus totale a breve termine (statico)* che viene creato dallo scambio economico mediato attraverso la piattaforma sia distribuito equamente tra la piattaforma e gli utenti business, ossia gli inserzionisti. Infatti, poiché la piattaforma, fintanto che riceve l'attenzione dei consumatori, gode di una posizione di forza economica nei confronti degli utenti commerciali affiliati, questa può essere potenzialmente abusata in vari modi: chiedendo un compenso eccessivo per l'intermediazione della piattaforma, richiedendo termini e condizioni non eque (ad esempio, per quanto riguarda l'accesso ai dati, la responsabilità, i tempi per le notifiche), attraverso un rifiuto ingiustificato e discriminatorio di accesso alla piattaforma, oppure ancora attraverso la discriminazione di terzi utenti commerciali al posto dei contenuti, servizi o prodotti propri delle piattaforme sulla stessa piattaforma [50].

In secondo luogo, il potere di mercato dei broker di attenzione, che determina un'eccessiva appropriazione di surplus degli utenti commerciali, può avere una serie di effetti preoccupanti anche dal punto di vista del *benessere dinamico del consumatore*. I broker di attenzione con un forte potere di mercato potrebbero avere un incentivo a creare una sorta di collo di bottiglia tra utenti business e potenziali clienti e ridurre l'offerta di pubblicità mirata. Questa riduzione permette alla piattaforma di ottenere margini più elevati ma, in condizioni standard, nel settore dei prodotti al dettaglio danneggia i consumatori, i quali devono fare i conti con prezzi più alti e una minore varietà dei prodotti.

3.4.3 Il modello di Prat e Valletti (2018)

Prat e Valletti, nel loro paper "*Attention Oligopoly*", esaminano le implicazioni della concentrazione del mercato dell'attenzione sui mercati dei prodotti e quindi sul benessere nel lungo termine dei consumatori finali.

Il punto di partenza è il seguente: l'attenzione dei consumatori è un input essenziale per i nuovi operatori nei mercati dei prodotti in quanto hanno bisogno di rendere i consumatori consapevoli della loro esistenza. Di conseguenza, è chiaro che i mercati dell'attenzione hanno un ruolo nell'influenzare l'ingresso e la concorrenza nei mercati dei prodotti. In particolare, può accadere che gli operatori storici (*incumbent*) precludano l'ingresso sul mercato pagando per l'attenzione di un gran numero di utenti.

Prat e Valletti dimostrano che il costo e l'efficacia di tale strategia di preclusione messa in atto dall'operatore storico, e quindi, in ultima analisi, la sua redditività, dipendono dal numero di intermediari dell'attenzione concorrenti che forniscono l'attenzione dei singoli utenti. Quanto più il mercato dell'attenzione è concentrato, tanto più tale pratica di preclusione nel mercato del prodotto è attraente, poiché la pubblicità totale si riduce e il numero di esborsi per garantire la preclusione diminuisce.

Da qui si spiega anche come le fusioni tra i broker di attenzione, e dunque il conseguente aumento della concentrazione, potrebbero avere un effetto importante sui consumatori attraverso l'aumento dei prezzi sui mercati dei prodotti (tra le altre cose Prat e Valletti, infatti, dimostrano pure che in tal caso l'effetto negativo dipende non solo dalla concentrazione ma dal grado di sovrapposizione dell'utilizzo. In particolare, nel caso in cui i due intermediari dell'attenzione non hanno alcun utente in comune, non ci sarebbe alcun effetto).

Inoltre, il lavoro di Prat e Valletti fornisce una giustificazione razionale anche di fenomeni apparentemente sconcertanti che vedono grandi società già affermate (ad esempio Amazon) fare pubblicità per parole chiave che contengono il proprio marchio nonostante i link organici appaiano già di per sé in cima ai risultati di ricerca. Questo comportamento che apparentemente sembra uno spreco di denaro ha in realtà una ragione ben precisa: evitare che un potenziale entrante "*rubi affari*" all'*incumbent* ed evitare la grande perdita di click (18-42%) che questi si troverebbe ad affrontare se un concorrente si impossessasse di quell'annuncio [49].

3.4.2.1 Ipotesi del modello

Al fine di mantenere l'esposizione semplice, il nucleo del paper di Prat e Valletti utilizza un modello "giocattolo", ossia impiega per comodità analitica delle ipotesi semplicistiche che permettono di caratterizzare in un modo relativamente semplice l'equilibrio in forma chiusa.¹⁶ Tali ipotesi sono le seguenti:

- Innanzitutto, ogni settore del commercio al dettaglio k è composto da un *incumbent* e un *entrante*, ciascuno dei quali può mostrare al massimo un annuncio. Inoltre, gli annunci sono considerati perfettamente informativi, ossia tramite la loro visione, il consumatore diventa perfettamente informato sul prodotto pubblicizzato.
- In secondo luogo, si presume che ogni piattaforma venda il proprio annuncio attraverso un'asta di secondo prezzo¹⁷, meccanismo solitamente utilizzato dalle piattaforme ad-supported.
- In terzo luogo, si presume che le piattaforme tengano le loro aste in sequenza, che per ogni utente l'ordine delle aste sia del tutto casuale e che vi siano informazioni complete sull'intera sequenza delle aste. L'entrante e l'incumbent partecipano ad ogni asta e il vincitore ha la possibilità di mostrare un annuncio a quell'utente.

3.4.2.2 Analisi del modello

Siano M l'insieme delle piattaforme digitali e J un sottoinsieme di esse utilizzate da un gruppo di utenti m_j (ovviamente $\sum_j m_j = 1$).

Queste piattaforme hanno informazioni proprietarie sulle preferenze dei loro utenti e vendono spazi pubblicitari mirati alle aziende (alle quali ci si riferisce in generale come *retail product industry*) che forniscono il prodotto o il servizio a cui l'utente è interessato. Ogni settore retail k è composto da un incumbent e un nuovo entrante che si contendono l'accesso al mercato dell'attenzione.

Gli acquirenti interessati al prodotto k hanno familiarità con il prodotto dell'operatore storico, ma non sono a conoscenza del prodotto del nuovo operatore. L'unico modo tramite cui possono conoscere quest'ultimo e colmare tale lacuna informativa è attraverso la visione del suo annuncio su una qualsiasi piattaforma di attenzione.

In particolare, per ogni singolo potenziale acquirente, si hanno due possibili condizioni:

¹⁶ L'ultima sezione del paper, tuttavia, rilassa ognuna di esse e mostra che la condizione di prelazione può essere estesa.

¹⁷ Il meccanismo dell'asta di secondo prezzo (o second price auction) è un meccanismo tale per cui chi si aggiudica l'asta non paga il prezzo offerto, bensì un centesimo in più di quanto offerto dal secondo miglior offerente. Supponendo che vi siano due inserzionisti, A e B, e che il primo offra 2€ e il secondo 1€, se l'inserzionista A si aggiudica l'asta non pagherà 2€ ogni qual volta un utente cliccherà sul suo annuncio ma 2,01€. Questo meccanismo tende ad indurre gli inserzionisti ad offrire l'importo più alto che sono disposti pagare, nella consapevolezza che spesso finiranno per pagare meno di tale importo.

- 1) L'acquirente è e resta a conoscenza solo del prodotto/servizio dell'incumbent ed in tal caso si avrà che:
 - u_1 : payoff atteso dell'acquirente (payoff lordo meno il prezzo);
 - π_1^I : payoff atteso dell'Incumbent (prezzo meno il costo di produzione) da quell'acquirente;
 - $\pi_1^E = 0$: payoff atteso dell'Entrante (prezzo meno il costo di produzione) da quell'acquirente.

- 2) L'acquirente viene a conoscenza del prodotto del nuovo operatore, ed in tal caso si avrà:
 - u_2 : payoff atteso dell'acquirente (payoff lordo meno il prezzo);
 - π_2^I : payoff atteso dell'Incumbent (prezzo meno il costo di produzione) da quell'acquirente;
 - π_2^E : payoff atteso dell'Entrante (prezzo meno il costo di produzione) da quell'acquirente.

Si presume che l'ingresso vada a beneficio dei consumatori, per cui il surplus del consumatore è maggiore se è a conoscenza del prodotto del nuovo operatore ($u_1 < u_2$), ma ciò riduce le entrate dell'operatore storico ($\pi_1^I > \pi_2^I$).

Ogni piattaforma gestisce un'asta di secondo prezzo per ciascuno dei suoi utenti, per cui per ogni utente ci saranno $n_j = \#J$ aste.

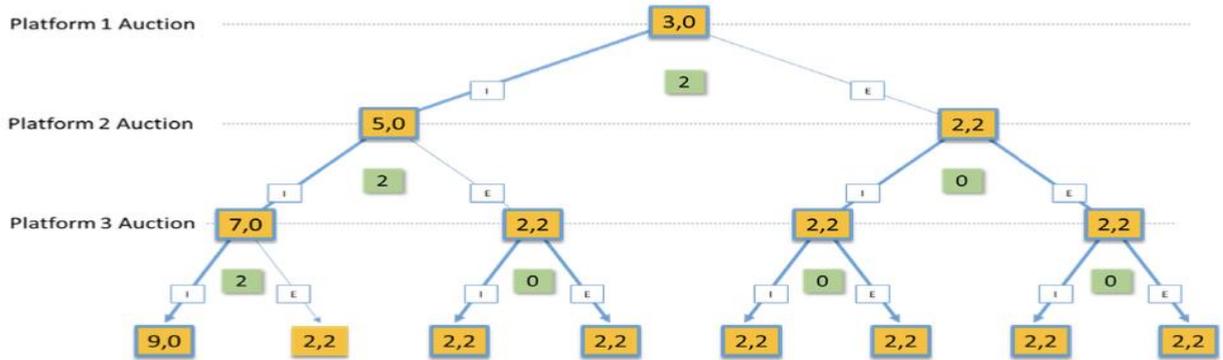
Dato un particolare settore k , si parte caratterizzando l'equilibrio di un'asta in un segmento J composto da n_j piattaforme. All'equilibrio si rappresenteranno con:

- R : i ricavi attesi della piattaforma in un segmento con n_j piattaforme, ovvero i ricavi previsti di una piattaforma prima che venga deciso l'ordine di vendita all'asta. Poiché le piattaforme sono simmetriche, questo è dato dai ricavi totali divisi per n_j .
- Π_I : il payoff netto dell'operatore storico in un segmento con n_j piattaforme, cioè il suo previsto payoff lordo, a seconda dell'equilibrio, meno quello che paga alle piattaforme.
- Π_E : il payoff netto del concorrente in un segmento con n_j piattaforme, ovvero il suo previsto payoff lordo, a seconda dell'equilibrio, meno quello che paga alle piattaforme.
- U : il payoff dei consumatori dato da u_1 o u_2 a seconda dell'equilibrio.

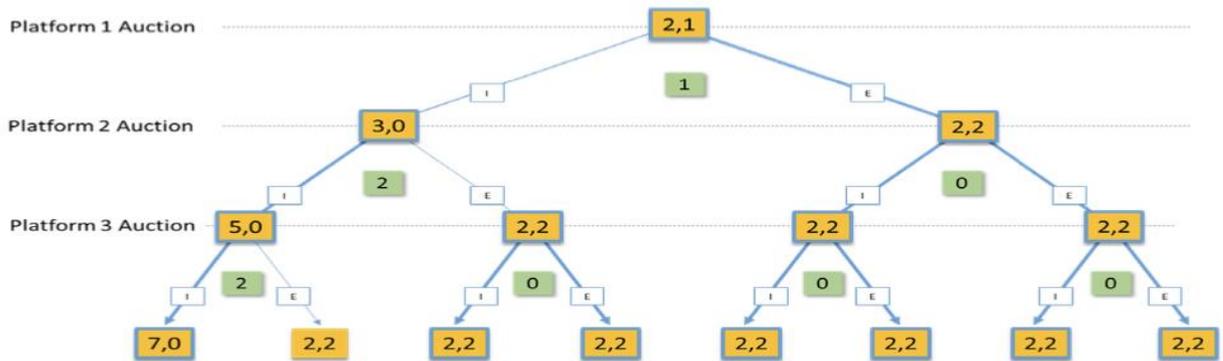
A titolo esemplificativo si considerino $n_j = 3$ piattaforme di proprietà indipendente e tre condizioni diverse come rappresentate in figura 3.8. In ognuno dei tre esempi il payoff lordo che l'entrante si attende dall'entrata sul mercato è pari a $\pi_2^E = 2$, così come pure il profitto lordo che si attende l'incumbent che finisce in duopolio, $\pi_2^I = 2$. Al contrario, il profitto di un operatore storico che riesce a difendere il suo monopolio varia nei tre casi: $\pi_1^I = 9$

nell'esempio 1, $\pi_1^I = 7$ nell'esempio 2 e $\pi_1^I = 5$ nell'esempio 3. Ad ogni condizione corrisponde un albero di gioco.

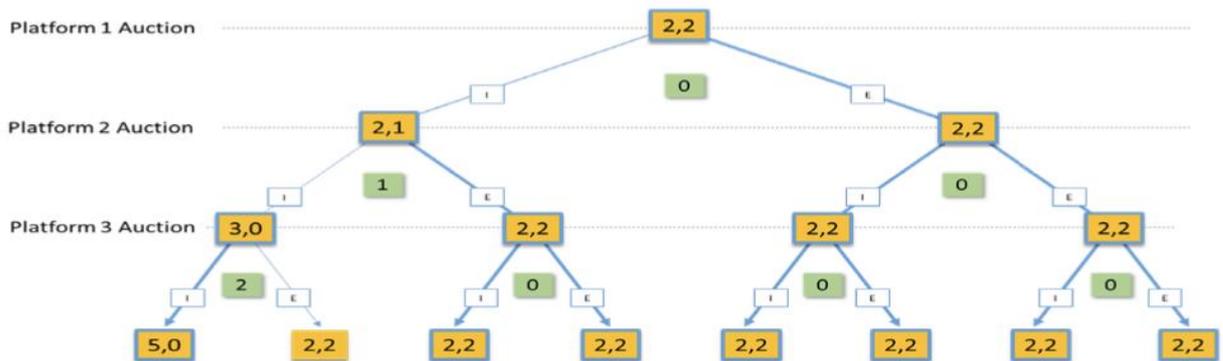
Example 1: $\pi_1 = 9, \pi_2 = \pi_E = 2 \rightarrow$ Incumbent Monopoly



Example 2: $\pi_1 = 7, \pi_2 = \pi_E = 2 \rightarrow$ Entry with Positive Platform Profit



Example 3: $\pi_1 = 5, \pi_2 = \pi_E = 2 \rightarrow$ Entry with Zero Platform Profit



- 3,0 Incumbent/Entrant continuation payoffs (net of payments in current and upcoming auctions)
- 2 Equilibrium bid in auction
- I E Auction winner (Incumbent/Entrant)

Figura 3.8: Alberi di gioco per la concorrenza tra un'azienda incumbent e un'azienda entrant per l'acquisto di un annuncio pubblicitario [49].

Ognuno dei nodi corrisponde ad un'asta. Come indicato in legenda i numeri nel riquadro in giallo rappresentano i payoff di continuazione di equilibrio a partire dal nodo superiore (ad esempio la coppia (5,0) appena sotto e a sinistra di (3,0) è il payoff di continuazione dopo che la prima asta è stata vinta dall'incumbent), mentre i numeri sotto ogni nodo rappresentano l'offerta di equilibrio per quell'asta. I rami di equilibrio sono indicati in grassetto. Si noti che una volta che l'incumbent ha perso un'asta, entrambi i giocatori sono indifferenti per quanto riguarda gli annunci rimanenti, quindi tutti i rami sono possibili percorsi di equilibrio.

Esempio 1: $\pi_1^I = 9, \pi_2^I = \pi_2^E = 2$

In questo scenario l'incumbent ha molto da guadagnare dalla difesa del monopolio per cui farà di tutto per scongiurare l'ingresso del concorrente. A tal fine sarà sufficiente che l'incumbent offra esattamente 2, che corrisponde alla massima offerta che potrebbe presentare il concorrente dato che in caso di successo questi si attende di guadagnare $\pi_2^E = 2$. Poiché dalla difesa del monopolio otterrebbe un payoff netto di $\Pi_I = \pi_1^I - n_J \pi_2^E = 3$, superiore di quello che otterrebbe in caso di duopolio ($\pi_2^I = 2$), all'equilibrio l'incumbent offrirà un 2 in tutte e tre le aste. In tal modo, l'incumbent vincerà tutte e tre le aste e otterrà $\Pi_I = \pi_1^I - n_J \pi_2^E = 3$, ogni piattaforma riceverà un ricavo di $\frac{n_J \pi_2^E}{n_J} = \pi_2^E$, l'entrante non riuscirà ad entrare e quindi otterrà un profitto nullo e il consumatore avrà un'utilità di u_1 . In questo caso si può osservare che l'operatore storico si troverà a pagare, per difendere il proprio monopolio, una parte sostanziale del suo payoff alle piattaforme.

Esempio 2: $\pi_1^I = 7, \pi_2^I = \pi_2^E = 2$

In questo caso, invece, l'operatore storico ha un po' meno da guadagnare dalla difesa del suo monopolio. Per escludere l'entrante in ognuna delle 3 piattaforme dovrebbe offrire per ciascuna asta $\pi_2^E = 2$ e si troverebbe alla fine del gioco con un payoff netto pari a $\pi_1^I - n_J \pi_2^E = 1$ che è inferiore a quello che otterrebbe nel caso in cui non riuscisse a difendere il monopolio ($\pi_2^I = 2$). In particolare, dalla difesa del monopolio, l'incumbent guadagnerebbe $\pi_1^I - \pi_2^I = 5$, per cui al massimo sarà disposto ad offrire 5 complessivamente nelle 3 aste. Dunque, l'entrante alla prima asta ne offre 1 e vince questo annuncio. L'offerta di 1 esattamente compensa l'importo che l'incumbent è disposto a pagare per vincere la prima asta in quanto prevede di dover sborsare un totale di 4 nelle due aste successive. In equilibrio, l'entrante ottiene un annuncio. In questo caso ciascuna piattaforma ottiene un ricavo pari a $\frac{\pi_1^I - \pi_2^I - (n_J - 1) \pi_2^E}{n_J} = \frac{1}{3}$, l'entrante e l'incumbent otterranno rispettivamente $n_J \pi_2^E - (\pi_1^I - \pi_2^I)$ e π_2^I , e il consumatore potrà beneficiare di un'utilità $u_2 > u_1$.

Esempio 3: $\pi_1^I = 5, \pi_2^I = \pi_2^E = 2$

In questo ultimo caso infine l'operatore storico ha ancora meno da guadagnare dalla difesa del suo monopolio, da cui guadagnerebbe $\pi_1^I - \pi_2^I = 3$. Di conseguenza, al massimo, sarà disposto ad offrire 3 complessivamente nelle tre aste. Dunque, l'entrante alla prima asta non offre nulla in cambio, in quanto questo esattamente compensa l'importo che l'incumbent è disposto a pagare per vincere la prima asta dato che prevede di dover sborsare un totale di 3 nelle due aste successive. In equilibrio quindi l'entrante ottiene un annuncio senza offrire nulla in cambio, ognuna delle tre piattaforme avrà un ricavo nullo e il consumatore avrà un'utilità $u_2 > u_1$.

Nello specifico, si tratta di un gioco sequenziale (dinamico) che si risolve con il metodo dell'*induzione a ritroso* (o *backward induction*). In particolare, si ha che dopo che tutte le n aste vengono eseguite, i payoff lordi dell'operatore storico e del concorrente sono rispettivamente $(\pi_1^I, 0)$ se l'operatore storico vince tutte le aste e (π_2^I, π_2^E) se il concorrente vince almeno un'asta.

Dopo $n - 1$ aste, sia l'operatore storico che il concorrente sono disposti a fare un'offerta al massimo pari a zero se l'entrante ha già vinto almeno un'asta, e sono disposti a fare un'offerta $(\pi_1^I - \pi_2^I, \pi_2^E)$ rispettivamente se l'operatore storico ha vinto tutte le aste precedenti, nel qual caso l'operatore storico vince l' n -esima asta se $\pi_1^I - \pi_2^E > \pi_2^E$.

I payoffs di continuazione per i due giocatori dopo $n - 1$ aste sono dunque:

$$(\Pi_I^{(n-1)}, \Pi_E^{(n-1)}) = \begin{cases} (\pi_1^I - \pi_2^E, 0) & \text{se l'incumbent ha vinto tutte le aste precedenti e } \pi_1^I - \pi_2^E > \pi_2^E \\ (\pi_2^I, \pi_2^E) & \text{altrimenti} \end{cases}$$

Per induzione si osserva che il payoff di continuazione dopo j aste è sempre (π_2^I, π_2^E) se l'entrante ha vinto almeno un'asta:

$$(\Pi_I^{(j)}, \Pi_E^{(j)}) = \begin{cases} (\Pi_I^{(j+1)} - \pi_2^E, 0) & \text{se l'incumbent ha vinto tutte le aste precedenti e } \pi_1^I - (n_j - j)\pi_2^E > \pi_2^E \\ (\pi_2^I, \pi_2^E) & \text{altrimenti} \end{cases}$$

All'inizio del gioco (quando $j = 0$), quindi, l'espressione di cui sopra diventa:

$$(\Pi_I, \Pi_E) = \begin{cases} (\pi_1^I - n\pi_2^E, 0) & \text{se } \pi_1^I - \pi_2^I > n\pi_2^E \\ (\pi_2^I, \pi_2^E) & \text{altrimenti} \end{cases}$$

Perciò, se $\pi_1^I - \pi_2^I > n\pi_2^E$, l'operatore storico vince tutte le aste e il totale dei ricavi della piattaforma è $n\pi_2^E$, mentre se $\pi_1^I - \pi_2^I < n\pi_2^E$ l'entrante vince la prima asta e paga $\max(\pi_1^I - \pi_2^I - (n - 1)\pi_2^E, 0)$ per il primo annuncio; in seguito tutte le offerte sono pari a zero e non importa chi riceve l'annuncio.

In definitiva, Prat e Valletti (2018) derivano la seguente proposizione (1) [49]:

Sia \bar{n} la soglia di monopolio dell'operatore storico definita come:

$$\bar{n} = \frac{\pi_1^I - \pi_2^I}{\pi_2^E} \quad (1)$$

- (a) Se $n_j < \bar{n}$, si ha un **monopolio dell'incumbent**: tutti gli annunci sono venduti al monopolista e ogni piattaforma ha un reddito π_2^E .
- (b) Se $\bar{n} + 1 > n_j > \bar{n}$, si ha un'**entrata con un profitto positivo**: almeno un annuncio viene venduto al nuovo concorrente e le entrate previste della piattaforma sono strettamente positive.
- (c) Se $n_j > \bar{n} + 1$, si ha un'**entrata con profitto zero**: almeno un annuncio viene venduto al nuovo concorrente e i ricavi previsti della piattaforma sono pari a zero.

I payoffs nei tre casi sono riportati nella tabella 3.1 seguente

Tabella 3.1: Condizioni di equilibrio delle aste nei tre casi [49]

		(a)	(b)	(c)
<i>Piattaforma</i>	R	π_2^E	$\frac{\pi_1^I - \pi_2^I - (n_j - 1) \pi_2^E}{n_j}$	0
<i>Incumbent</i>	Π_I	$\pi_1^I - n_j \pi_2^E$	π_2^I	π_2^I
<i>Entrante</i>	Π_E	0	$n_j \pi_2^E - (\pi_1^I - \pi_2^I)$	π_2^E
<i>Consumatore</i>	U	u_1	u_2	u_2

Ne deriva che la probabilità che il consumatore venga a conoscenza del prodotto/servizio del nuovo operatore aumenta con il numero di piattaforme di proprietà indipendente utilizzate da quel consumatore. Inoltre, tanto meno il settore delle piattaforme dell'attenzione è concentrato tanto più sarà difficile per l'operatore storico precludere l'accesso al concorrente su tutte le piattaforme disponibili.

Una volta caratterizzato ciò che accade all'interno di ogni segmento di consumo, si rivolge l'attenzione all'intero settore delle piattaforme J , considerando il seguente timing:

1. In ogni settore k , appare un potenziale entrante, il cui profitto lordo per utente in caso di ingresso è π_2^E , una variabile casuale ricavata da una distribuzione cumulativa F (non correlata tra i vari settori).
2. In ogni segmento J , viene selezionato un ordine casuale delle piattaforme in J e viene eseguita una sequenza di aste secondo tale ordine.
3. In ogni segmento J , i payoffs per i consumatori, le piattaforme, l'operatore storico e l'entrante derivano dall'esito dell'asta.

Il sotto-gioco che inizia nel secondo punto, per ogni segmento J , è caratterizzato dall'equilibrio definito prima nella proposizione (1).

Infatti, si ha che se e solo se $n_j > \frac{\pi_1^I - \pi_2^I}{\pi_2^E}$ l'entrante riuscirà ad acquistare un annuncio nel segmento J , e ciò avviene con una probabilità pari a $1 - F\left(\frac{\pi_1^I - \pi_2^I}{n_j}\right)$. Aggregando quindi per i segmenti di consumo si ottiene che il surplus atteso dei consumatori è dato da:

$$\bar{U} = u_1 - (u_2 - u_1) \sum_J m_J F\left(\frac{\pi_1^I - \pi_2^I}{n_j}\right)$$

Il fatto che un particolare segmento J rimanga un monopolio dipende quindi dal confronto tra il profitto che l'entrante si attende di realizzare, π_2^E , e il numero di piattaforme n_j utilizzate in quel segmento.

Nell'ipotesi che il profitto atteso sia uniformemente distribuito e che dunque F sia una distribuzione uniforme con supporto $[0, M]$, l'espressione del surplus atteso dei consumatori diventa particolarmente semplice:

$$\bar{U} = a - b \sum_J m_J \frac{1}{n_j}$$

Dove $a = u_2$ e $b = \frac{(u_1 - u_2)(\pi_1^I - \pi_2^I)}{M} > 0$ sono costanti.

Ne discende che un aumento della concentrazione - rappresentata dal termine $\sum_J m_J \frac{1}{n_j}$ quale media ponderata della quota di piattaforma tra i segmenti - porta ad un aumento della probabilità che l'incumbent riesca a precludere all'entrante l'ingresso in uno o più segmenti e conseguentemente ad una perdita di benessere per il consumatore.

3.4.2.3 Conclusioni del modello

In conclusione, il lavoro di Prat e Valletti esplora un interessante meccanismo in base al quale le piattaforme di attenzione potrebbero cercare di sfruttare il potere di mercato minacciando di fornire una via di accesso al mercato per i nuovi operatori ed estrarre il surplus dai monopoli a monte, ossia nell'industria dei prodotti.

Si analizza l'effetto della concorrenza tra le piattaforme di attenzione attraverso l'impatto che essa ha su mercati dei prodotti e se ne deriva una condizione tale per cui è come se il potere di mercato generasse a sua volta potere di mercato. In questo senso, l'effetto di una maggiore concentrazione delle piattaforme di attenzione non è positivo né per i nuovi operatori innovativi (che hanno bisogno che gli annunci pubblicitari siano conosciuti dagli utenti), né per gli utenti (che finiscono per acquistare a prezzi elevati dai produttori storici).

CAPITOLO 4

Big Data: tra concorrenza e privacy

Il panorama digitale online ha permesso di beneficiare di ondate di innovazione che hanno trasformato radicalmente il modo in cui le persone cercano informazioni, acquistano prodotti o servizi e comunicano tra di loro. Che sia su un computer, un tablet o un qualsiasi dispositivo mobile connesso è come se qualsiasi cosa fosse letteralmente a portata di mano. In questo scenario, non è possibile compiere una qualsiasi attività online senza che questa lasci una traccia di dati come una sorta di impronta digitale.

Contemporaneamente, come osservato nei capitoli precedenti, l'estrazione di valore dai dati personali degli utenti è diventata una fonte significativa di potere per i maggiori operatori dei mercati online. L'emergere di *data-opoli*¹⁸, ossia di un gruppo di società che controllano attraverso le loro piattaforme un volume significativo e una varietà di flussi di dati personali incomparabile per altre imprese, ha posto l'attenzione non solo sui problemi di concorrenza che possono nascere ma anche relativamente ai rischi in termini di privacy che il trattamento di tali dati e la concentrazione in mano di pochi attori potrebbero comportare.

Sebbene infatti, il trattamento dei dati personali non debba essere demonizzato di per sé, in quanto le aziende possono utilizzare le informazioni sugli utenti per adattare un servizio alle loro specifiche esigenze, esiste un delicato compromesso tra la privacy dell'individuo a cui le informazioni si riferiscono e la divulgazione di dati al fine di un servizio migliore, personalizzato e a volte anche "gratuito".

Pertanto, nell'era digitale odierna in cui il trattamento dei dati personali è un'attività critica per la competitività tra le imprese, la dimensione della privacy e della concorrenza sono profondamente collegate. Innanzitutto, perché, soprattutto in presenza di servizi offerti ad un prezzo monetario nullo, la privacy e la protezione dei dati potrebbero essere un elemento distintivo di qualità su cui le imprese possono competere. A tal proposito ci si chiede se i mercati digitali soddisfino le esigenze di privacy dei singoli individui o se ci possono essere gravi problemi di fallimento del mercato che richiedono rimedi normativi.

Il dubbio che la debole concorrenza possa portare ad un'eccessiva raccolta di dati privati e ad un'insufficiente offerta di opzioni di privacy per soddisfare le diverse preferenze degli utenti è, infatti, una delle questioni che le posizioni dominanti di alcune società operanti online, tra cui Google e Facebook, hanno sollevato negli ultimi anni. Si potrebbe affermare, ad esempio, che un mercato ben funzionante offrirebbe opzioni di privacy molto più diverse, vale a dire soluzioni più specifiche di opt-in e opt-out¹⁹ per la raccolta dei dati, nonché

¹⁸ Si fa riferimento in particolare alle quattro società statunitensi Google, Apple, Facebook e Amazon (o GAF4 in breve) per cui l'Economist ha anche creato l'acronimo BAADD ("*too Big, Anti-competitive, Addictive and Destructive to democracy*"). <https://www.economist.com/briefing/2018/01/20/the-techlash-against-amazon-facebook-and-google-and-what-they-can-do>.

¹⁹ Con i termini opt-in e opt-out (in cui *opt* è l'abbreviazione di *option*: opzione) si fa riferimento alle opzioni che rispettivamente forniscono o derogano il consenso alla raccolta e al trattamento dei propri dati personali.

un'opzione per l'utilizzo dei servizi offerti da questi giganti digitali anche con un canone mensile di abbonamento senza la raccolta di dati e/o pubblicità [52]. In secondo luogo, perché le norme e i regolamenti che disciplinano la raccolta e l'accesso ai dati dei consumatori potrebbero avere, in parallelo, implicazioni sulla dimensione economica e sulla concorrenza delle aziende.

4.1 Privacy e sua tutela nei mercati digitali

La privacy è un concetto difficile e complesso, interpretabile come una delimitazione tra la sfera privata e quella pubblica di un individuo, legata quindi alla protezione dello spazio personale e spesso resa operativa per via del diritto alla salvaguardia e al controllo delle informazioni personali. Nella trasposizione digitale delle nostre vite, essa assume un significato nuovo, costituisce un potenziale elemento distintivo della qualità dei servizi online e pone delle sfide inedite che richiedono un'attenta considerazione.

4.1.1 Atteggiamento dei consumatori: il “paradosso della privacy”

La comprensione degli atteggiamenti e dei comportamenti dei consumatori in relazione alla privacy e alla protezione dei dati è una questione chiave nell'analisi delle dinamiche competitive nei mercati che coinvolgono i dati degli utenti.

In linea generale numerosi sondaggi hanno rilevato che i consumatori apprezzano la privacy e sono inoltre sempre più preoccupati per la sua tutela online. Un'indagine condotta nel 2015 dalla Commissione Europea su oltre 15.000 intervistati ha rilevato che più di otto su dieci ritenevano di non avere un controllo completo dei loro dati personali che fornivano online e che i due terzi erano preoccupati di ciò [53].

Tuttavia, esistono una serie di fattori che possono rendere difficile comprendere l'*importanza effettiva* per i consumatori della privacy e della protezione dei dati, primo fra tutti il cosiddetto “*paradosso della privacy*”, ossia l'evidenza empirica di una discrepanza tra le notevoli preoccupazioni che i consumatori esprimono relativamente alla privacy e il loro effettivo comportamento online che mostra come questi spesso non siano cauti nel divulgare informazioni private.

Le possibili spiegazioni dell'esistenza di tale paradosso sono tra le più varie. Innanzitutto, occorre dire che le preferenze in materia di privacy sono eterogenee e di conseguenza esiste una soggettività per quanto riguarda gli atteggiamenti a riguardo [54]. In secondo luogo, anche per lo stesso individuo, la scelta se condividere o meno le proprie informazioni personali risulta essere influenzata dal contesto e dalle circostanze specifiche in cui le informazioni sono richieste [54].

Inoltre, pure i pregiudizi comportamentali, rivestono una grande importanza in relazione alle decisioni in materia di privacy. A tal proposito, particolarmente problematica potrebbe essere la miopia dei consumatori che condividono i propri dati in vista di un beneficio immediato senza considerare le implicazioni sulla privacy a lungo termine di queste scelte. La presenza di asimmetrie informative tra i consumatori e i fornitori di servizi online solleva poi ulteriori preoccupazioni circa la capacità dei consumatori di comprendere e agire in

merito alle informative sulla privacy. I consumatori possono avere, infatti, informazioni imperfette laddove sia troppo costoso o difficile per loro impegnarsi in modo efficace con tutte le informazioni rilevanti per le loro decisioni. Uno dei motivi potrebbe essere ricondotto alla lunghezza e alla complessità dei termini informativi. Dall'indagine precedentemente citata, risulta infatti che solo il 20% degli intervistati dichiarava di aver letto integralmente le dichiarazioni sulla privacy [53]. A tal proposito anche l'ACCC nella sua *"Digital Platforms Inquiry"* conferma che: *"... pochi consumatori sono pienamente informati, comprendono appieno, o controllano efficacemente, la portata dei dati raccolti e l'accordo che stanno concludendo con le piattaforme digitali quando si iscrivono o utilizzano i loro servizi"* [51]. Sempre dalla stessa indagine è emerso che il 36% dei consumatori ritiene, erroneamente, che la mera esistenza di una politica sulla privacy significhi che le imprese non condivideranno informazioni personali con terzi [51].

In definitiva, la mancanza di informazioni, condizioni lunghe e complesse e pregiudizi comportamentali, possono condurre a scenari in cui i consumatori condividono involontariamente più dati di quanto farebbero se prendessero decisioni pienamente informate e imparziali. Tutti questi fattori, presi nel loro insieme, potrebbero dunque:

- (i) rendere difficile comprendere in quali circostanze e mercati i consumatori comprendono effettivamente e valorizzano la privacy, sia in teoria che in pratica e
- (ii) compromettere l'efficacia delle leggi sulla protezione dei dati che si basano principalmente sul consenso dei consumatori per promuovere buoni risultati in materia di protezione dei dati [55].

4.1.2 La privacy nell'economia "zero-price"

Come già osservato, nei mercati digitali in cui è centrale la raccolta e l'utilizzo dei dati dei consumatori a fini commerciali, non è infrequente l'offerta di prodotti o servizi a prezzo zero. Nel paragrafo 3.2.2 del capitolo precedente è stato anche dimostrato come questa pratica non sia un'espressione di un pricing predatorio, ma al contrario, nella logica dei mercati multi-versante, possa essere il risultato, anche per i monopolisti, di un comportamento di pricing ottimale.

Tuttavia, tale condizione pone certamente una serie di difficoltà, di natura sia pratica che concettuale, nell'individuare i problemi di concorrenza in questi mercati. Ad esempio, da un punto di vista normativo un prezzo nullo complica la definizione di un mercato rilevante, che costituisce un prerequisito al fine di constatare un'eventuale posizione dominante e dunque la possibilità per un'azienda, pur applicando appunto un prezzo nullo, di influenzare unilateralmente le condizioni di scambio con i consumatori a suo vantaggio e a danno di questi ultimi.

Infatti, uno strumento comunemente usato per definire i confini del mercato rilevante è il cosiddetto *test del "piccolo ma significativo aumento non transitorio del prezzo"* o in breve *test SSNIP* (Small but Significant Non-transitory Increase in Price), il quale definisce il mercato rilevante come il più piccolo insieme di prodotti sostitutivi in modo tale che un

aumento di prezzo sostanziale (di solito il 5% o il 10%) e non transitorio (di solito un anno) da parte di un ipotetico monopolista sarebbe redditizio²⁰ [56].

Tuttavia, poiché questo test si basa essenzialmente sui meccanismi di prezzo, la presenza di prodotti o servizi offerti a prezzo nullo rende la sua applicazione non possibile. Di conseguenza, in tali circostanze, come soluzione alternativa per la definizione del mercato rilevante, è stata proposta quella di considerare, anziché gli aumenti di prezzo, le riduzioni di qualità che un operatore con potere di mercato potrebbe implementare, e di usare quindi il *test SSNDQ* (Small but Significant Non-transitory Decrease in Quality) per misurare l'effetto di una "*piccola ma significativa diminuzione non transitoria della qualità*" [18].

Ecco dunque che in questo scenario la privacy può essere intesa come una potenziale dimensione di qualità nei mercati a prezzo zero e come una misura degli effetti della condotta delle imprese. Per cui nel contesto di un mercato a prezzo zero la definizione di posizione dominante quale capacità di peggiorare unilateralmente la qualità attraverso una riduzione della privacy, potrebbe essere concettualmente equivalente alla definizione di posizione dominante in un mercato a prezzo positivo quale capacità di fissare in modo permanente prezzi elevati.

Quantificare però il grado di protezione della privacy garantito da una data impresa è tutt'altro che semplice, in quanto non esistono a riguardo delle misure oggettivamente condivise. A tal proposito si potrebbero usare sondaggi tra i consumatori per misurare la fiducia che questi hanno nelle disposizioni sulla privacy di un'impresa, anche se queste percezioni, come osservato nella sezione precedente, potrebbero non corrispondere alla realtà. Così come si potrebbe usare come misura di privacy la categorizzazione della quantità dei dati raccolti sulla base (1) dell'ambito di raccolta dei dati, ossia in sostanza del numero di variabili per le quali i dati sono raccolti, (2) della frequenza di raccolta dei dati (ad esempio, se sono raccolti solo nel momento in cui un consumatore crea un account o ogni volta che utilizza il servizio), (3) se la raccolta dei dati è limitata all'interazione del consumatore con il servizio in questione o se continua mentre il consumatore utilizza altri servizi, e (4) del grado di condivisione di questi dati con altre parti, sia interne che esterne all'impresa. Tuttavia, per essere informativo come misura di privacy, il volume dei dati raccolti deve essere messo in relazione al modo in cui vengono processati e trattati dall'azienda, così come le garanzie di sicurezza dei dati dell'azienda [58].

Pertanto, occorre infine dire che, in assenza di misure della qualità in termini di protezione della privacy e dei dati dei consumatori, appropriate, ben accettate e quantificabili obiettivamente, non è ancora chiaro come questo test possa essere reso operativo nella pratica. Proprio per questo l'OCSE afferma che si tratta di un'idea sì utile come guida concettuale ma non di uno strumento preciso che i tribunali e le autorità garanti della concorrenza dovrebbero effettivamente tentare di applicare [40].

²⁰ Il test SSNIP rende operativa l'idea del test del monopolista ipotetico, (Hypothetical Monopolist Test – HMT), secondo il quale (almeno) un'impresa con una quota di mercato del 100%, cioè un monopolista, deve essere in grado di esercitare un potere di mercato nel mercato considerato. Se ciò non fosse vero, secondo l'HMT, il mercato rilevante è stato definito in modo troppo ampio e deve essere definito in modo più restrittivo [57].

4.1.3 Regolare la privacy online: il GDPR

Considerate le difficoltà per i consumatori di gestire la privacy online in mezzo a trade-off sempre più complessi, negli anni si è reso necessario pensare, al posto della responsabilità individuale, ad un intervento regolamentare al fine di bilanciare gli interessi dei consumatori contro il potere delle entità commerciali che detengono i loro dati.

È in questo contesto che è stato concepito e sviluppato dall'Unione Europea il Regolamento UE 2016/679, ossia il *Regolamento Generale sulla Protezione dei Dati*, noto anche con la locuzione *GDPR* (General Data Protection Regulation) [59].

Nato per rafforzare e armonizzare la regolamentazione in materia di privacy e di raccolta e utilizzo dei dati personali all'interno dell'UE, e considerato ormai la pietra angolare della legge europea sulla privacy, il GDPR è entrato in vigore dal 24 maggio 2016, ma è stato reso efficace solo a partire dal 25 maggio 2018 per consentire alle parti interessate di adeguarsi.

Tale regolamento disciplina in sostanza il trattamento dei dati personali²¹ dei residenti nell'UE ed è vincolante sia per le imprese europee che per le imprese estere.

L'imposizione delle sue disposizioni non solo alle imprese con sede in Europa, ma a qualsiasi entità che tratta i dati personali dei cittadini europei, indipendentemente dalla sede centrale e dal luogo in cui viene effettuato il trattamento (art. 3 del GDPR), costituisce sicuramente uno degli elementi più innovativi rispetto alla precedente Direttiva 95/46/CE che ha sostituito e su cui si basa. Questa scelta è stata presa alla luce della facilità con cui i dati sono scambiati a livello globale che dà origine a due preoccupazioni: in primo luogo, potrebbe esserci un rischio per i consumatori europei di assoggettare il trattamento dei propri dati a norme straniere relative alla privacy meno rigorose; in secondo luogo, l'essere soggette a regole più severe, potrebbe comportare, proprio per via di questa condizione di disparità, uno svantaggio competitivo per le aziende digitali europee rispetto ai loro concorrenti stranieri [60].

Le figure chiave, titolari di diritti e doveri, considerate dal regolamento sono (art. 4 GDPR):

- Il **data subject**, ossia l'*interessato al trattamento* che corrisponde alla persona fisica a cui i dati personali si riferiscono;
- Il **data controller**, ossia il *titolare del trattamento*, la persona fisica o giuridica, l'autorità pubblica, il servizio o altro organismo che determina le finalità, le modalità e gli strumenti del trattamento dei dati personali.

²¹ Ai sensi dell' articolo 4 del GDPR per «dato personale» si intende “qualsiasi informazione riguardante una persona fisica identificata o identificabile («interessato»); si considera identificabile la persona fisica che può essere identificata, direttamente o indirettamente, con particolare riferimento a un identificativo come il nome, un numero di identificazione, dati relativi all'ubicazione, un identificativo online o a uno o più elementi caratteristici della sua identità fisica, fisiologica, genetica, psichica, economica, culturale o sociale”. Di conseguenza i dati resi anonimi non sono soggetti al regime di diritti e doveri stabilito dal GDPR (si consideri il Considerando 26 del GDPR), anche se diversi data scientists hanno dimostrato che l'anonimizzazione dei dati personali a livello individuale in modo tale che gli individui non possano essere nuovamente identificati è molto difficile [40].

- Il *data processor*, ovvero il *responsabile del trattamento*, la persona fisica, giuridica, autorità pubblica o qualsiasi ente che elabora i dati personali per conto del titolare del trattamento.

4.1.3.1 I principi fondamentali del GDPR

La disciplina del GDPR, improntata a disciplinare in maniera più efficace la crescente elaborazione digitale dei dati personali, si fonda su sette principi chiave [59]:

1. *Principio di liceità, correttezza e trasparenza* (art 5, n.1, a) secondo cui è previsto che i dati personali dell'interessato siano trattati in modo lecito, corretto e trasparente. In particolare, il trattamento è lecito laddove l'interessato abbia espresso il proprio consenso informato al trattamento dei dati per una o più specifiche finalità, laddove sia derivato dall'esecuzione di un contratto, da un obbligo legale o da interessi legittimi del data controller. Per ciò che concerne la trasparenza, essa, oltre ad essere un diritto dell'interessato, costituisce un obbligo a carico del titolare e del responsabile del trattamento i quali sono tenuti a fornire al data subject informazioni e comunicazioni relative al trattamento, tra cui l'identità del titolare, le finalità ed i diritti dell'interessato, in modo tale che queste siano facilmente accessibili e comprensibili [61];
2. *Principio di limitazione delle finalità* (art 5, n.1, b) secondo cui i dati personali devono essere raccolti per finalità specifiche, esplicite e legittime. Non è vietato l'ulteriore trattamento per una finalità secondaria, purché questa non sia "incompatibile" con le finalità per le quali i dati sono stati raccolti.
3. *Principio di minimizzazione dei dati* (art 5, n.1, c) secondo il quale i dati personali soggetti a trattamento devono essere adeguati, pertinenti e limitati a quanto è necessario per gli scopi per cui sono stati raccolti e trattati.
4. *Esattezza dei dati* (art 5, n.1, d) che richiede che i dati siano sempre esatti e aggiornati, per cui di conseguenza eventuali inesattezze devono essere tempestivamente rettificate o cancellate.
5. *Principio di limitazione della conservazione* (art 5, n.1, e) che impone che i dati siano conservati in una forma che consenta l'identificazione degli interessati per un periodo di tempo non superiore a quello necessario al perseguimento delle finalità per le quali i dati personali sono trattati.
6. *Principio di integrità e riservatezza* (art 5, n.1, f) per cui il trattamento dei dati deve essere eseguito in modo da garantirne un'adeguata sicurezza attraverso l'adozione di misure di sicurezza tecniche ed organizzative adeguate a proteggere i dati stessi da trattamenti non autorizzati o illeciti, dalla loro perdita o distruzione o dal danno accidentale.
7. *Principio di responsabilità* (art 5, n.2) per cui il responsabile del trattamento non solo è responsabile del rispetto dei precedenti principi ma è anche in grado di comprovarne l'osservanza.

4.1.3.2 Il ruolo dell'interessato al trattamento

Il GDPR riconosce un empowerment del soggetto interessato al trattamento dei dati in quanto impone quale condizione essenziale per il trattamento dei dati il suo consenso.

In particolare, l'articolo 4 del regolamento precisa che per consenso si debba intendere una manifestazione *“libera, specifica, informata e inequivocabile, della volontà, dell'interessato”* data attraverso una dichiarazione o tramite una chiara azione affermativa, con la quale lo stesso manifesta il proprio assenso al trattamento dei suoi dati personali. Ne consegue che le forme tacite di consenso, quali ad esempio il silenzio o le caselle preselezionate non sono valide.

Inoltre, affinché un consenso sia considerato valido si sostiene che debba essere *“liberamente dato”*. È chiaro dunque che, al fine della valutazione della validità del consenso, l'esistenza di un eventuale potere di mercato dovrebbe essere importante. Laddove infatti una grande azienda goda di una posizione dominante, il singolo interessato potrebbe non avere scelta di fatto scelta nell'assenza di alternative [40].

4.1.3.3. Il diritto alla portabilità dei dati

Uno dei diritti che il GDPR riconosce all'interessato al trattamento, che è potenzialmente molto rilevante ai fini della concorrenza, è *il diritto alla portabilità dei dati* (art. 20 GDPR) che consente all'interessato di ricevere i propri dati personali dal titolare del trattamento a cui li ha forniti, in un formato *“strutturato, di uso comune e leggibile da dispositivo automatico”* e di poterli trasmettere ad un altro titolare. Secondo tale diritto un consumatore che ha ad esempio pubblicato dei contenuti o messo dei likes su un social network ha la possibilità di trasferire facilmente questi suoi dati su un social network concorrente.

La motivazione per la quale questo diritto è stato introdotto è stata quella di fornire all'interessato una sorta di grado di protezione contro gli effetti lock-ins indotti dai dati che, come ben noto, costituiscono una barriera all'ingresso per le nuove imprese e conseguentemente un potenziale fattore di rischio per la concorrenza [60]. Inoltre, ciò potrebbe avere anche il potenziale di ridurre il vantaggio accumulato negli anni da aziende leader nella raccolta e nel trattamento di dati personali.

Tuttavia, è ancora poco chiaro il modo in cui questo diritto debba essere interpretato e implementato. Innanzitutto, mentre è certo che nel perimetro dei dati *“portabili”* rientrano i dati volontari e ne sono invece esclusi quelli dedotti dal titolare del trattamento, è poco chiaro se anche i dati osservati sono coperti dall'art. 20, par.1 del GDPR²² [61].

Poiché questi ultimi sono molto più difficili per un nuovo concorrente, rispetto ai dati volontari, sia da raccogliere direttamente che da acquistare da intermediari, la loro inclusione o meno nella categoria dei dati portabili è chiaro che potrebbe avere delle significative implicazioni sulla concorrenza.

²² Secondo le linee guida europee predisposte da parte del Working Party 29, un gruppo di esperti europei della privacy, tra i dati portabili rientrerebbero anche i dati osservati, ossia i dati relativi all'osservazione dell'attività dell'utente nell'interazione con il titolare, di parere contrario invece risulta essere la Commissione Europea che interpreta la norma in senso restrittivo considerando come trasferibili solo i dati forniti volontariamente dall'interessato [61].

Inoltre, alcuni analisti hanno espresso il timore che la portabilità dei dati possa essere costosa, per cui se imposta in modo indistinto a società piccole e startup così come a imprese dominanti, potrebbe finire per imporre oneri eccessivi alle prime rispetto alle seconde, ottenendo così l'effetto opposto a quello desiderato [60].

4.1.3.4 Responsabilità congiunta del titolare e del responsabile del trattamento

Parallelamente all'empowerment del soggetto interessato, il GDPR rafforza anche la responsabilità di chi esegue il trattamento. Infatti, a carico del titolare e del responsabile del trattamento vi sono una serie di obblighi primo fra tutti quello di trasparenza, il quale impone che le informazioni fornite all'interessato siano rese con un linguaggio semplice e chiaro, abbiano sempre forma scritta ed includano le finalità del trattamento, il periodo di conservazione dei dati personali, l'intenzione di trasferire i dati a un destinatario in un paese terzo (art 14 GDPR).

È importante, inoltre, porre l'accento sul principio della responsabilità congiunta del titolare e del responsabile del trattamento. Secondo tale principio, i titolari del trattamento, come ad esempio i siti web che esternalizzano compiti quali l'analisi dei dati a fornitori di tecnologia terzi, non possono esimersi facilmente dalla responsabilità per le violazioni della privacy che si verificano mentre collaborano con il fornitore di tecnologia che riveste il ruolo di responsabile del trattamento. Nello specifico, infatti, il titolare del trattamento è tenuto ad informare i propri utenti in merito a qualsiasi trasferimento di dati personali a fornitori terzi (art. 13, n. 1, e) del GDPR) e può essere corresponsabile con lui per la violazione delle disposizioni del GDPR.

Infine, per i responsabili e i titolari del trattamento rilevante è anche l'inasprimento delle sanzioni in caso di violazione degli obblighi previsti dal regolamento. In particolare, le sanzioni possono arrivare, a seconda di quale dei due importi sia superiore, fino a 20 milioni di euro ovvero fino al 4% del fatturato annuo di un'impresa a livello mondiale (art. 83, par. 5 e 6 del GDPR).

4.2 Impatto del GDPR sulla concorrenza

Se è indubbio che dall'attuale quadro normativo del GDPR i cittadini europei ne escano rafforzati in termini di tutela della privacy e dei dati personali, sono meno chiari gli effetti indiretti sul loro benessere in virtù dell'impatto del nuovo regolamento sulla competitività delle imprese che si occupano del trattamento di dati personali.

Il GDPR, infatti, come osservato nel paragrafo precedente, impone a carico delle imprese che trattano i dati personali dei cittadini europei una serie di oneri in modo indistinto, ossia sia che esse siano imprese europee o straniere e sia che siano piccole imprese o multinazionali. Poiché quindi le imprese devono sostenere dei costi al fine di conformarsi alla legislazione sulla privacy e sulla protezione dei dati, vi è il rischio, laddove tali costi siano in gran parte fissi, che ciò svantaggi in modo sproporzionato le imprese più piccole e che incidano sulla concorrenza in un modo quanto meno indesiderato. A riguardo, l'OECD stima che i costi di conformità al GDPR siano sostanziali [62]. Ad esempio, 1,7 milioni di

sterline per impresa è la media dei costi stimata nel Regno Unito, con un range che varia da poco meno di 1 milione di sterline per le imprese con 100-249 dipendenti e 2,3 milioni di sterline per le imprese con oltre 1.000 dipendenti. Negli Stati Uniti, invece, un sondaggio condotto dalla società di consulenza PwC [63] stima che la spesa per conformarsi al GDPR per il 68% delle aziende statunitensi varia tra 1 milione e 10 milioni di dollari. In Europa, per le PMI si stimano dei costi annuali di IT per conformarsi al GDPR che rappresentano tra il 17- 40% dei budget IT annuali pre-GDPR [62].

Si teme, inoltre, che il GDPR, poiché fonda la regolamentazione della privacy sul consenso dato dall'utente, possa avvantaggiare e radicare gli operatori storici più grandi, soprattutto quelli che operano su più mercati per i quali è indubbiamente più facile ottenere l'assenso da parte degli utenti.

Diversi gli studi che negli ultimi anni hanno cercato di comprendere le implicazioni delle riforme del GDPR recentemente attuate. Tra questi, lo studio di Batikas et al. (2020), presentato di seguito, dimostra come il regolamento abbia avuto l'effetto indesiderato di aumentare la concentrazione nel mercato dei servizi di tecnologia web di tracking di terze parti, di cui ormai si serve la maggior parte dei servizi online.

4.2.1 Effetti indesiderati del GDPR: Studio empirico Batikas et al. (2020)

Questo studio dimostra come la legge sulla privacy interagisce con la concorrenza nel contesto del GDPR. In particolare, i suoi risultati empirici documentano un aumento della concentrazione del mercato tra i fornitori di tecnologia web dopo l'introduzione del regolamento europeo. Di seguito sono riportati la strategia empirica, i dati ed i risultati conseguiti in tale studio.

4.2.1.1 Contesto dell'analisi

I siti web spesso utilizzano servizi di tecnologia di tracking di terze parti per il monitoraggio comportamentale e la profilazione del loro pubblico. Molte delle interazioni tra l'editore del sito web e questi fornitori di tecnologia di terze parti avvengono in tempo reale ogni qual volta un'utente visita il sito web dell'editore e si possono osservare attraverso le richieste HTTP che il sito web fa durante il caricamento dei contenuti.

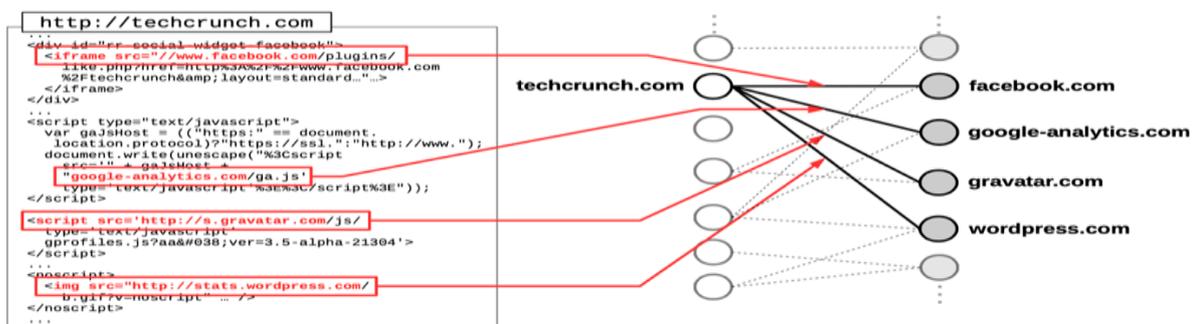


Figura 4.1: Esempio di richieste di terze parti [64]

La figura 4.1 ne illustra un esempio stilizzato: la pagina web ospitata all'indirizzo techcrunch.com effettua richieste a domini di terze parti, quali google-analytics.com

e **facebook.com**, incorporando così contenuti ospitati su server di terze parti, come script e immagini, i quali possono differire per ogni richiesta e per ogni utente. Esempi di contenuti incorporati sulla pagina tramite questi fornitori possono essere la pubblicità ma anche funzionalità interattive come i sistemi di commento o di chat, quali i pulsanti “mi piace”, “condividi” e “login”.

Poiché questi servizi richiamati in tempo reale mentre un utente naviga su un sito web, sono in grado di (i) risalire all’indirizzo IP dell’utente che ha originato la richiesta HTTP, (ii) rispondere alla richiesta inviando un cookie da memorizzare sul suo dispositivo e (iii) monitorare quindi l’attività dell’utente durante la navigazione, si tratta di tecnologie di tracking *privacy-intrusive*, ossia invasive della privacy. Pertanto, secondo la normativa del GDPR il sito web che utilizza queste tecnologie - come cookies, script, immagini, web trackers o tecnologie di fingerprinting - deve raccogliere il consenso degli utenti di cui tratta i dati personali ed è responsabile in solido con i fornitori di tecnologia terzi che eseguono la raccolta e l’analisi di tali dati per conto del titolare del trattamento, cioè del sito web.

Il settore dei servizi di tecnologia web risulta essere un settore notevolmente concentrato, con i primi 20 servizi che coprono circa un terzo del mercato [64]. In particolare, dopo l’introduzione del GDPR il panorama dei servizi di tracking sul web pare abbia subito dei cambiamenti significativi. Innanzitutto, si osserva una tendenza generale al ribasso del numero di tracker nei siti web che si rivolgono ad un pubblico UE e sono quindi soggetti direttamente a normativa, mentre succede il contrario a quelli che si rivolgono ad un pubblico non UE e non sono quindi soggetti al GDPR. Tale dinamica si può osservare nella figura 4.2 derivante da uno studio²³ condotto su 2000 siti web tra cui alcuni che si rivolgono ad un pubblico UE ed altri ad un pubblico non UE, nello specifico statunitense.

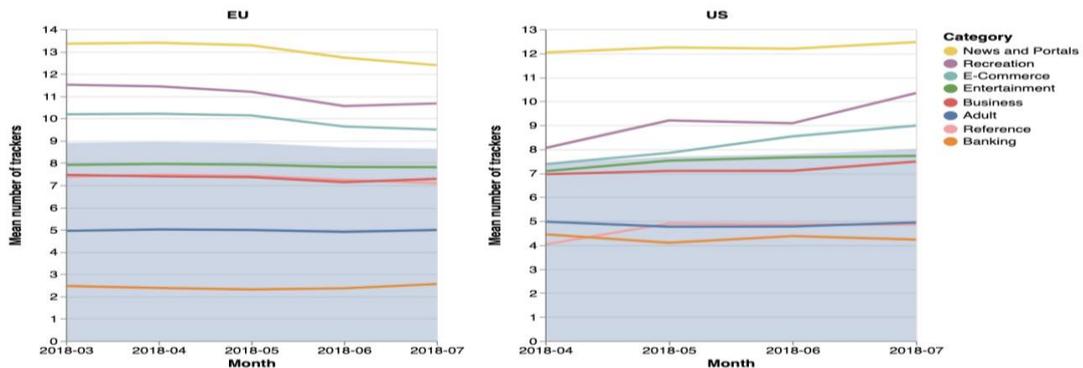


Figura 4.2: Andamento numero medio di tracker di terze parti sui siti web con pubblico UE e US. [65]

In secondo luogo, soprattutto per quanto riguarda i fornitori di tecnologia per la pubblicità online si può osservare, come si evince dalla figura 4.3, anche qui una maggiore concentrazione per i siti web con pubblico UE, con in particolare due società, Google e Facebook che in seguito all’entrata in vigore del GDPR hanno mantenuto rispetto agli altri provider sostanzialmente invariata o di poco ridotta la loro quota di mercato.

²³ Studio condotto da whotracks.me: <https://whotracks.me/blog/gdpr-what-happened.html#fn:4>

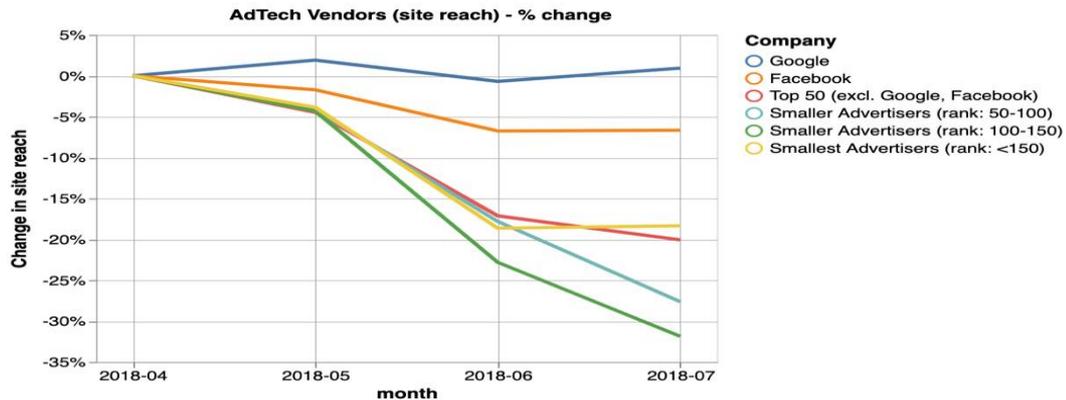


Figura 4.3: Andamento site reach, ossia della % dei siti web in cui il tracker è presente [65].

Utilizzando la site reach di un tracker come proxy della quota di mercato, si può ritenere che il GDPR potrebbe aver avuto effetti regressivi sulla concorrenza nello spazio pubblicitario online in Europa [65].

La domanda a cui Batikas et al. (2020) cercano di rispondere con il loro studio è quindi la seguente: il GDPR, progettato per migliorare la privacy degli utenti nel web, ha avuto effetti negativi sulla concorrenza?

4.2.1.2 Dati dell'analisi

Il data-set utilizzato proviene dalla combinazione di fonti di dati pubblici e proprietari che permette di considerare un campione che copre un arco temporale che interessa sia un periodo pre-GDPR (1 maggio 2017 – 24 maggio 2018) che un periodo post-GDPR (25 maggio 2018 – 15 novembre 2018).

In particolare, il set di dati principali fa riferimento all'*HTTArchive* di Google BigQuery, che periodicamente fa il crawl delle homepage di circa mezzo milione di siti web host. Tranne poche eccezioni²⁴, questi dati sono disponibili su base bisettimanale, derivanti da un crawl all'inizio e a metà del mese. Tramite questo data-set lo studio definisce un campione di 110.706 siti web host osservati nei 33 snapshot di dati disponibili, ossia 21 punti di dati prima dell'introduzione del GDPR e 12 punti di dati dopo, per un totale di 3.653.298 osservazioni. Per ogni sito web host sono state raccolte informazioni sull'identità delle terze parti a cui l'host invia richieste http e il numero totale di richieste di terze parti.

Per stabilire se uno specifico sito web host si rivolge ad un pubblico dell'UE si fa riferimento al suo dominio di primo livello (TLD=top level domain) specifico del paese come .it per l'Italia, .de per la Germania o .fr per la Francia.

Inoltre, dal database di *whotracks.me* si ricava una categorizzazione dei domini di terze parti in base ai diversi tipi di servizi (ad esempio pubblicità, social media, analisi) e una corrispondenza tra i domini di terze parti e le aziende che li gestiscono. Ad esempio, Google gestisce una varietà di domini diversi, compresi quelli meno ovvi come doubleclick.net, invitemediamedia.com e 2mdn.net.

²⁴ Il data-set pubblico disponibile su Google's BigQuery web service, non contiene informazioni per il 12-01-2016, 01-01-2017, 01-15-2017, 04-15-2018, e 05-01-2018.

4.2.1.3 Modello econometrico di base

Si considera la seguente specificazione di base:

$$(1) \text{Log}(\text{Number3rdParties}_{it} + 1) = \alpha + \beta_1 \text{Trend}_t + \beta_2 (\text{Trend}_t \times \text{EU-Audience}_i) + \\ + \delta_1 \text{Post}_t + \delta_2 (\text{Post}_t \times \text{EU-Audience}_i) + \mu_i + \varepsilon_{it}$$

Dove:

- $\text{Number3rdParties}_{it}$ è il numero di domini di terze parti che il sito web i richiede al momento t .
- Post_t indica il periodo successivo all'entrata in vigore del GDPR il 25 maggio 2018.
- Trend_t rappresenta gli andamenti temporali lineari specifici del gruppo ed un trend espresso in giorni, dove 1 maggio 2017 corrisponde a 0,01.
- EU-Audience_i corrisponde ad una variabile booleana che indica se il sito web i ha un dominio di primo livello specifico di un paese dell'UE, quindi se si rivolge ad un pubblico UE, ed è conseguentemente soggetto alle norme del GDPR.
- μ_i gli effetti fissi specifici del sito web i , che dovrebbero catturare le configurazioni organizzative invariante nel tempo non osservate e specifiche dell'azienda, nonché le preferenze di rischio e di conformità.
- ε_{it} il termine di errore, caratterizzato dalle ipotesi standard

Questo modello permette di effettuare un confronto prima/dopo all'interno del gruppo di siti web che possono avere incentivi relativamente piccoli a reagire al GDPR (δ_1 , cioè non UE), e un confronto prima/dopo all'interno del gruppo di siti web che possono avere incentivi relativamente grandi a reagire al GDPR ($\delta_1 + \delta_2$, cioè UE), nonché un confronto di tali cambiamenti tra i vari gruppi (δ_2).

Negli output presenti nelle tabelle seguenti sono rappresentati tra parentesi gli errori standard per ciascun stimatore, mentre i diversi valori di p -value sono rappresentati dalla seguente notazione: * per $p < 0,10$, ** per $p < 0,05$, *** per $p < 0,01$.

4.2.1.4 Richieste a domini di terzi-parti

Lo studio cerca di analizzare inizialmente se a causa del GDPR i siti web riducono il numero di richieste ai fornitori terzi di tecnologia di tracking.

Dalla figura 4.4, che mostra il numero medio (di log) di richieste di terze parti per ogni sito web dell'UE e non UE, si può osservare come subito dopo l'introduzione del GDPR vi sia stato un calo sostanziale ed improvviso del numero di domini di terzi richiesti per i siti web con pubblico UE. Si rileva anche una riduzione per i siti web con pubblico extra UE ma di minore entità.

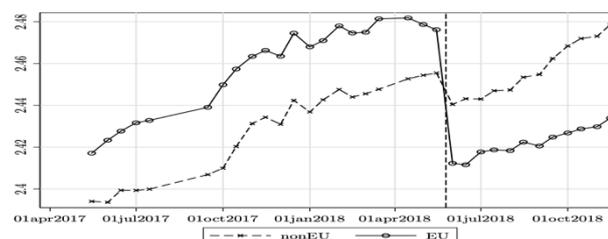


Figura 4.4: Log del numero di richieste a domini di terze parti [64]

Considerando l'output della regressione (1), riportato in tabella 4.1, in cui appunto si considera come variabile di interesse il numero dei domini di terze parti che il sito web-host richiede e si controlla per gli effetti fissi dei siti web host e un andamento temporale lineare, si stima che questa riduzione sia del -8,1% (UE) e del -2,4% (non UE).

Tabella 4.1: *Cambiamento nel numero di richieste a domini di terze parti*

	(1) EU	(2) NonEU	(3) All
Trend	0.0173*** (0.0008)	0.0194*** (0.0004)	0.0194*** (0.0004)
Post	-0.0840*** (0.0027)	-0.0245*** (0.0011)	-0.0245*** (0.0011)
Trend × EU			-0.0021** (0.0009)
Post × EU			-0.0596*** (0.0029)
Observations	728112	2925186	3653298
$\overline{R^2}$	0.8903	0.9033	0.9007
Mean DV	2.4453	2.4385	2.4399

Nonostante queste riduzioni, lo studio ha rilevato un effetto del GDPR in tal senso poco sostenibile, in quanto, come si evince dalla figura 4.5 seguente, che rappresenta le previsioni²⁵ di un modello con tendenze temporali specifiche per gruppo e periodo, i siti web con un pubblico non UE sono rimbalzati al livello delle richieste di dominio di terzi che esistevano direttamente prima del GDPR solo quattro mesi dopo il GDPR, mentre i siti web con un pubblico UE tornano al livello iniziale dopo 22 mesi.

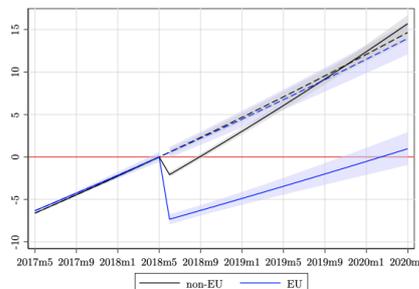


Figure 4.5: *Variazione del numero di domini di terze parti richiesti, previsione out-of-sample [64]*

Una possibile spiegazione di questo fenomeno può essere individuata nella crescente complessità del panorama di conformità che le aziende hanno dovuto affrontare dopo il GDPR, il quale inizialmente ha di fatto aumentato in modo significativo l'esposizione dei siti web a causa dell'incertezza giuridica. Questo probabilmente ha fatto sì che i set web subito dopo la sua entrata in vigore adottassero un approccio prudente nell'interpretazione del regolamento. Ad esempio, si può pensare che, dati i rischi creati dal regime di responsabilità congiunta del GDPR, laddove i siti web non erano in grado di valutare

²⁵ Nota: Previsioni basate su modelli di regressione che tengono conto dei diversi trend nei siti web UE/non UE e prima/dopo il GDPR, rispettivamente. Normalizzate al mese precedente l'introduzione del GDPR e in percentuale ($\exp(\text{coeff}) - 1 * 100$). Le linee tratteggiate indicano tendenze controfactuali, le aree ombreggiate indicano fasce di confidenza del 90%.

accuratamente i rischi per la privacy presso il loro fornitore di tecnologia web, il modo migliore per ridurre la loro esposizione è stato quello di ridurre l'uso di fornitori di tecnologia web, preferendo magari l'insourcing del monitoraggio comportamentale e la profilazione dei consumatori. Tuttavia, è ragionevole pensare che nel corso del tempo, con la pubblicazione di nuove linee guida politiche, questa incertezza è via via diminuita e i siti web, che hanno imparato ad adattarsi interpretando tali linee guida, cercando un consulente legale e osservando i concorrenti abbiano abbandonato la grande cautela iniziale e siano diventati di nuovo più attivi nel generare richieste da parte di terzi.

4.2.1.5 Struttura del mercato e concorrenza

Lo studio procede poi ad indagare se e come la struttura del mercato delle tecnologie web sia cambiata dopo il GDPR nei mercati UE e non UE.

Concentrazione nei mercati UE e non UE

A tal scopo, si considera come indice di concentrazione del mercato l'*Indice Herfindahl-Hirschman (HHI)* dato dalla somma delle quote di mercato dei fornitori di tecnologia web al quadrato:

$$HHI = \sum_{i=1}^N s_i^2$$

Dove $s_i = \frac{n_i}{\sum_{k=1}^N n_k}$ rappresenta la quota di mercato relativa del fornitore i data dalla portata (*reach*) del fornitore i , ossia il numero di siti web che utilizzano uno o più domini di terze parti di un dato fornitore di tecnologia web i diviso la portata totale di tutti i fornitori N nella categoria k .

Partendo da un'analisi grafica si può visualizzare dalle figure 4.6 seguenti come si sviluppa nel tempo l'indice in questione. I mercati sono definiti, in base al pubblico a cui si rivolgono i siti web, in mercati UE e non-UE.

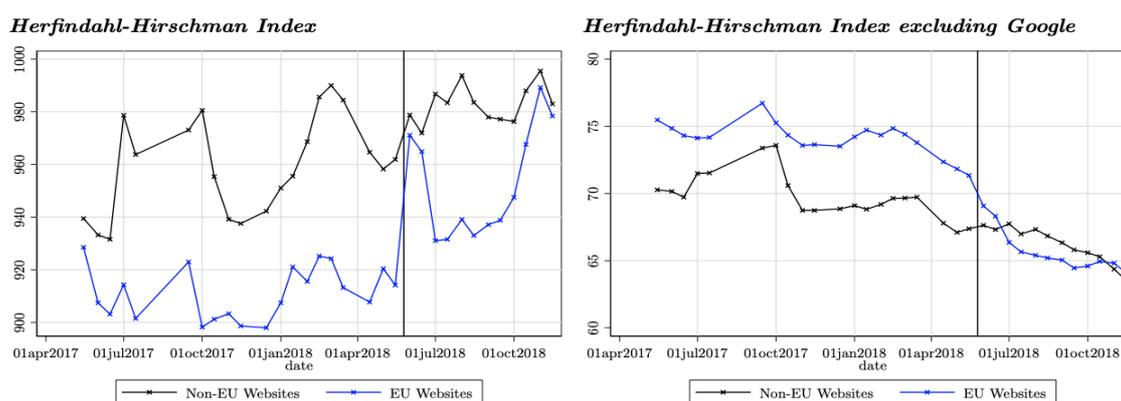


Figura 4.6: Cambiamento nella struttura di mercato del settore delle tecnologie di tracking web [64]

Mentre prima dell'entrata in vigore del GDPR (il 25 maggio 2018) per il mercato UE non si osserva un trend visibile nell'HHI, sembra esserci al contrario un grande aumento dopo la sua introduzione al punto tale che alla fine del periodo d'osservazione dello studio il mercato UE ha raggiunto quello non-UE.

Se si esclude dall'analisi grafica l'azienda con la più grande quota di mercato, che è Google sia nel mercato UE che in quello non UE, si può notare il suo evidente predominio, ancor più se si confrontano le scale dei due grafici. Infatti, l'HHI che include Google varia approssimativamente tra 900 e 1000, mentre la variazione nell'ipotetico HHI che esclude Google è solo approssimativamente tra 60 e 80. Inoltre, si osserva una tendenza al ribasso in entrambi i mercati, con una forte diminuzione nel mercato UE dopo l'entrata del GDPR

Lo studio cerca quindi di stimare questo effetto attraverso una regressione caratterizzata dalla seguente specificazione:

$$(2) \quad HHI = \alpha + \beta_1 Trend_t + \beta_2 (Trend_t \times EU-Audience_i + \delta_1 Post_t + \delta_2 (Post_t \times EU-Audience_i) + EU-Audience_i + \mu_i + \varepsilon_{it}$$

Di cui si osserva l'output nella tabella 4.2 seguente:

Tabella 4.2: Cambiamenti nella struttura di mercato del settore della tecnologia web

	All		w/o Google	
	(1)	(2)	(3)	(4)
Trend	6.3468** (2.8259)	3.3995 (2.7583)	-1.0873*** (0.2307)	-1.2842*** (0.2167)
EU	-42.5656*** (9.2568)	-37.6137*** (11.3974)	3.8699*** (0.7978)	3.9458*** (0.8206)
Post	5.7988 (8.1989)	11.8158 (8.3108)	-0.5481 (0.5472)	0.0669 (0.7122)
Trend × EUs	-3.0778 (3.7099)	-4.8635 (4.5255)	0.2351 (0.3074)	0.1575 (0.3047)
Post × EU	26.3429** (11.4205)	30.3990** (12.9623)	-5.5375*** (0.9579)	-5.2732*** (1.1626)
Observations	64	74	64	74
$\overline{R^2}$	0.7732	0.7009	0.9044	0.7977
Mean DV	947.5992	947.5509	69.6881	70.0832

Con un caveat reso necessario dalla piccola dimensione del campione che naturalmente discende dall'aggregazione a livello di mercato, queste regressioni suggeriscono che l'HHI nel mercato UE è aumentato significativamente di 26,3 (30,4) punti da una base di 911,3 (914,5) punti o 2,9% (3,3%). Mentre non si trova che la concentrazione nel mercato extra-UE sia aumentata in modo significativo.

Le stime puntuali di $Post_t$ sono anche molto più piccole delle stime puntuali di $Post_t \times EU-Audience_i$ web con sede nell'UE. Passando all'ipotetico HHI nelle colonne (3) e (4), si trova una diminuzione di 5,5 (5,3) punti da una base di 74,1 (74,2) punti o 7,4% (7,1%). Ciò suggerisce che l'aumento della concentrazione complessiva del mercato sembra essere spiegato esclusivamente da un aumento del mercato dell'impresa dominante.

Quote di mercato

Per comprendere se effettivamente questa concentrazione sia principalmente da imputare alla crescita della quota di mercato dell'impresa dominante si passa ad un'analisi di regressione in cui si considera come variabile di interesse la quota di mercato congiunta di tutti i tracker di un determinato fornitore di tecnologia, $s_i = \frac{n_i}{\sum_{k=1}^N n_k}$

Rispetto alla specificazione di base le variabili ulteriori sono: *Top2* che indica la quota di mercato congiunta di tutti i domini tracker delle due aziende con la quota di mercato più alta (Google e Facebook), *Google* che indica la quota di mercato congiunta di tutti i domini tracker dell'azienda, *No. Domains* e *No. Categories* che corrispondono rispettivamente al numero di domini ricollegabili ad un determinato fornitore e al numero di sotto-mercati²⁶ in cui lo stesso è attivo.

L'output delle regressioni è illustrato nella tabella 4.3 seguente in cui gli effetti fissi e gli errori standard sono raggruppati a livello di dominio tracker nelle colonne (1)-(2), a livello di impresa nelle colonne (3)-(6).

Tabella 4.3: Cambiamento nelle quote di mercato dei fornitori di tecnologia web

	(1)	(2)	(3)	(4)	(5)
Post	-0.000000 (0.000011)	-0.000773*** (0.000162)	-0.012304*** (0.002404)	-0.000001 (0.000011)	-0.000002 (0.000011)
Post × EU	0.000000 (0.000015)	-0.004263*** (0.000604)	-0.062390*** (0.014866)	-0.000004 (0.000011)	-0.000009 (0.000012)
EU × No. Domains		-0.006523*** (0.000592)			
Post × No. Domains		0.000766*** (0.000164)			
Post × EU × No. Domains		0.004228*** (0.000601)			
EU × No. Categories			-0.089749*** (0.022195)		
Post × No. Categories			0.012301*** (0.002404)		
Post × EU × No. Categories			0.062375*** (0.014866)		
EU × Top2				-0.215112 (0.408395)	
Post × Top2				0.035368 (0.048553)	
Post × EU × Top2				0.141851 (0.285693)	
EU × Google					-0.792670*** (0.000034)
Post × Google					0.104032*** (0.000011)
Post × EU × Google					0.545883*** (0.000012)
Observations	4050688	4050688	4050688	4050688	4050688
\bar{R}^2	0.997941	0.998188	0.998151	0.998001	0.998215
Mean DV	0.001580	0.001580	0.001580	0.001580	0.001580
% Google-Effect nonEU					0.346784
% Google-Effect EU					2.221910

Nella colonna (1) non si individuano cambiamenti nella quota di mercato dell'azienda media, né per quanto riguarda il mercato dei siti web con pubblico UE, né per quanto riguarda il mercato dei siti web con pubblico non UE.

Nella colonna (2), si verifica l'eterogeneità tra le aziende, utilizzando la variazione del numero di domini richiesti dai siti web del campione in esame, che si identificano come appartenenti ad un fornitore di tecnologia web²⁷. Dall'analisi si stima un effetto di

²⁶ Questi sotto-mercati sono definiti secondo la classificazione da *whotracks.me* ed includono *Advertising*, *Commenti*, *Content Delivery Network*, *Customer Interaction*, *Essential*, *Hosting*, *Varie*, *Pornvertising*, *Site Analytics*, *Social Media*, *Video/Audio Player* e *Unknown*. Per una descrizione dettagliata di queste categorie si veda https://whotracks.me/blog/tracker_categories.html.

²⁷ Lo studio rileva che ad esempio, i siti web nel campione richiedono i servizi di Facebook utilizzando 4 domini distinti, mentre nel caso di Google questo numero è pari a 110 (due terzi di tutte le richieste a Google vengono però effettuate tramite 5 domini). L'azienda media di tecnologia web ha 1,01 domini (deviazione standard 0,46) [64].

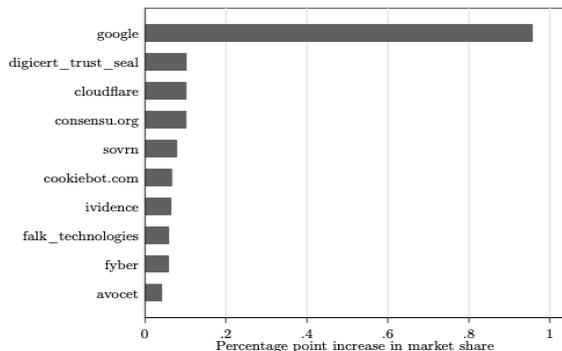
interazione lineare delle variabili *Post* e $Post \times EU$ con il numero di domini. Si può osservare una significativa diminuzione della quota di mercato media nel mercato non-UE, così come nel mercato UE (i coefficienti di *Post* e $Post \times EU$ sono negativi) dopo l'entrata in vigore del GDPR. Tuttavia, questa situazione è invertita per i fornitori di tecnologia web con almeno tre domini.

Si effettua dunque un'analisi simile nella colonna (3), utilizzando la variazione del numero di sotto-mercati in cui è attivo un fornitore di tecnologia web. In questo caso si osserva un'inversione della diminuzione media della quota di mercato per i fornitori di tecnologia web che sono attivi in almeno due sotto-mercati.

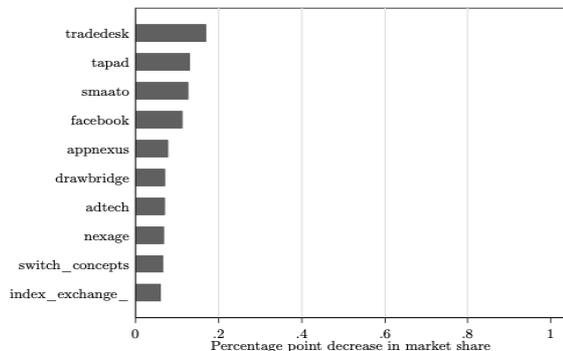
Pare inoltre che le imprese più grandi siano state in grado di aumentare la loro quota di mercato dopo l'entrata in vigore del GDPR, soprattutto nel mercato dell'UE, ma anche in quello non-UE. Per fare più chiarezza su questo aspetto, lo studio confronta in particolare i due fornitori di tecnologia web con la quota di mercato più grande (Facebook e Google) con tutte le altre aziende nella colonna (4). Si osserva che le stime di questo punto aumentano drasticamente. Tuttavia, i coefficienti stimati non sono significativamente diversi da zero. Nella colonna (5) invece si confronta l'azienda con la quota di mercato più grande (Google) con tutte le altre aziende e si osserva che la quota di mercato di Google è aumentata in modo significativo. Le stime suggeriscono un aumento di 0,1 punti percentuali per quanto riguarda il mercato non-UE e un aumento di 0,5 punti percentuali per quanto riguarda il mercato UE. Per quanto riguarda la quota di mercato pre-GDPR di Google con i siti web extra-UE e con quelli dell'UE, si tratta di un aumento rispettivamente dello 0,3% e del 2,2%.

Nelle figure seguenti si possono quindi osservare le variazioni delle quote di mercato²⁸ delle dieci imprese con i maggiori aumenti ("*winner*") e delle dieci imprese con i maggiori cali ("*losers*") nel mercato UE ed extra UE, due settimane (figura 4.7) prima e due settimane dopo e sei mesi (figura 4.8) prima e sei mesi dopo l'entrata in vigore del GDPR. Mentre le aziende che perdono la maggior parte delle quote di mercato variano a seconda degli orizzonti temporali e dei mercati, Google si distingue come il chiaro vincitore nel breve e nel lungo periodo, sia nel mercato UE che in quello non-UE.

Short-run winners in the EU market

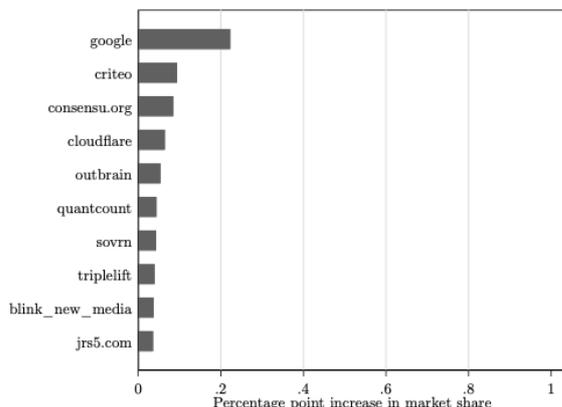


Short-run losers in the EU market



²⁸ Le quote di mercato sono definite come il numero di host di siti web con domini di primo livello specifici di un paese (non) dell'UE che inviano richieste a qualsiasi dominio di un'impresa rispetto al numero di siti web host con domini specifici di un paese (non) dell'UE che inviano richieste a domini di terzi parti.

Short-run winners in the non-EU market



Short-run losers in the non-EU market

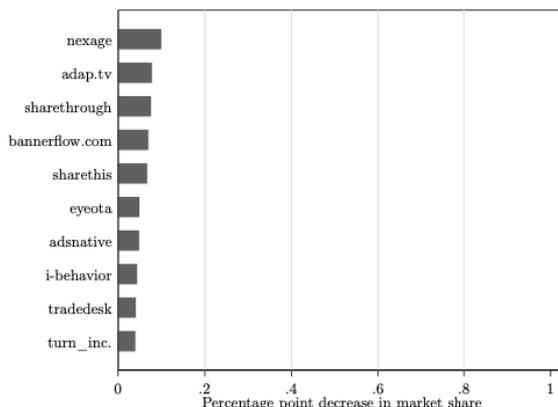
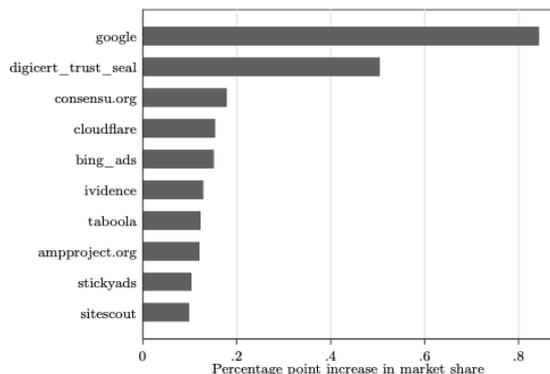
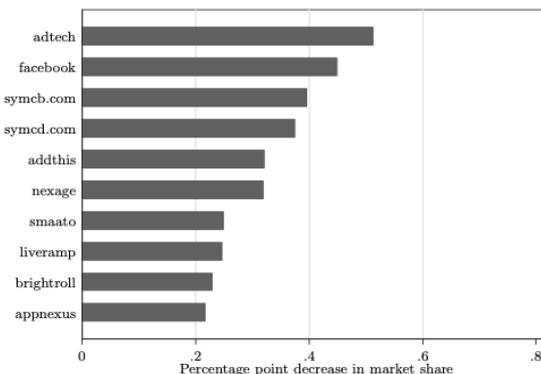


Figura 4.7: Winners e losers 2 settimane prima/dopo l'entrata in vigore del GDPR [64]

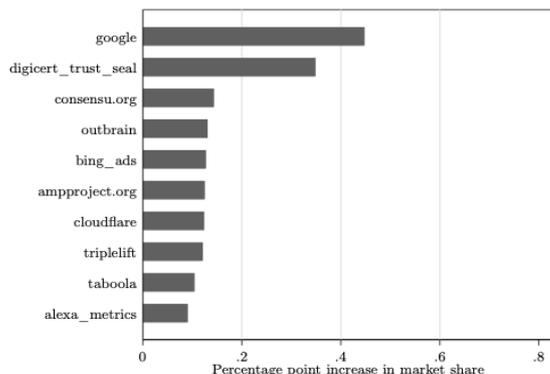
Winners in the EU market



Losers EU market



Winners in the non-EU market



Losers in the non-EU market

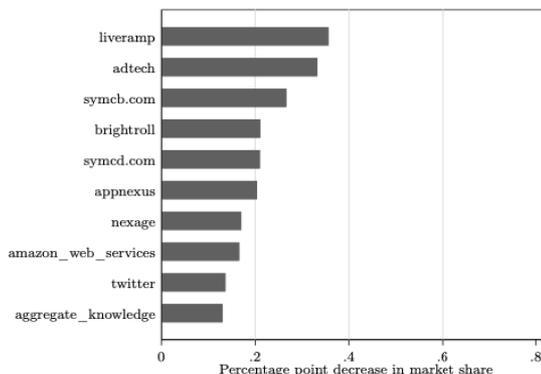


Figura 4.8: Winners e losers 2 settimane prima/dopo l'entrata in vigore del GDPR [64]

Per capire in quali categorie di servizi Google ha guadagnato quote di mercato dopo l'entrata in vigore del GDPR lo studio distingue ulteriormente tra i sotto-mercati in cui Google è attivo, ciascuno dei quali corrisponde ad una colonna nell'output raffigurato nella tabella 4.4, ed effettua una serie di regressioni in cui considera come variabile dipendente la quota di mercato di Google ossia $s_i = \frac{n_i}{\sum_{k=1}^N n_k}$ dove n_i è il numero di siti web che inviano richieste al tracker i di proprietà di Google, e in cui *Google* indica la quota di mercato congiunta di tutti i domini tracker dell'azienda nella categoria indicata dal titolo della rispettiva colonna.

Tabella 4.4: Cambiamenti nelle quote di mercato, per sottomercato, Google vs tutte le altre aziende

	(1) Video	(2) Advertising	(3) Tagmanager	(4) Analytics	(5) CDN/API	(6) Other
EU	1.5e-11 (.000134)	-4.2e-12 (.000061)	-2.8e-12 (.000165)	-3.5e-12 (.000087)	5.3e-11 (.000052)	-4.5e-12 (.000071)
Post	-0.000019 (.000044)	7.0e-06 (.000023)	-0.000012 (.000058)	-0.000012 (.000024)	-0.000015 (.000019)	9.2e-07 (.000032)
Post × Google	1.20566*** (.000031)	-0.441783*** (.00003)	.731151*** (.000035)	.78715*** (.000031)	.971121*** (.000022)	-.058424*** (.000038)
Post × EU	.000113 (.000085)	-0.000029 (.000029)	.000093 (.000102)	-0.000031 (.000035)	.000031 (.000031)	-1.5e-06 (.000032)
Post × EU × Google	-7.15514*** (.00006)	1.83314*** (.000042)	-5.91109*** (.000087)	1.9904*** (.000072)	-1.97809*** (.000028)	.095707*** (.000072)
Observations	4050688	4050688	4050688	4050688	4050688	4050688
\bar{R}^2	.998331	.995211	.99491	.993765	.999512	.90875
Mean DV	.00158	.00158	.00158	.00158	.00158	.00158
% Effect nonEU	1.59465	-1.58394	1.19014	1.95054	1.39394	-16.5879
% Effect EU	-7.46261	5.40282	-7.96724	7.23468	-1.39309	15.722

I risultati rafforzano quelli precedenti, in quanto si osservano cambiamenti significativi solo nelle quote di mercato di Google, e non per le altre aziende. In primo luogo, considerando le variazioni della quota di mercato di Google per quanto riguarda i siti web non-UE, l'aumento maggiore è 1,9% per quanto riguarda la categoria *Analytics*, mentre la categoria *Altri/Sconosciuti* subisce il calo maggiore con un -16,6%, categoria in cui lo studio indica però che già prima del GDPR la quota di mercato di Google era inferiore allo 0,25%.

Per quanto riguarda i siti web dell'Unione Europea, si osservano sostanziali diminuzioni nel mercato dei lettori audio/video player (-7,5%), nel mercato dei tag manager (-8,0%) e nel mercato dei CDN/API (-1,4%). I maggiori aumenti invece si osservano nel mercato degli *Analytics* (7,2%) e nel mercato pubblicitario (5,4%).

Si può quindi concludere che l'effetto medio dello 0,3% di aumento della quota di mercato complessiva di Google con i siti web extra-UE e l'effetto medio del 2,2% di aumento della quota di mercato complessiva di Google con i siti web UE osservato in tabella 4.3 nascondono una significativa eterogeneità. Almeno nel mercato dei siti web UE, Google si rafforza soprattutto nei mercati in cui era più forte prima del GDPR (pubblicità e analytics) dove vantava quote di mercato rispettivamente del 25,8% e del 38,4%.

Quote di mercato di Google e Facebook nell'Advertising

Rivolgendo infine l'attenzione esclusivamente al sottomercato della rete pubblicitaria, e considerando come variabile dipendente la quota di mercato del provider lo studio trova (vedi tabella 4.5) che Google e Facebook hanno entrambi aumentato la quota di mercato nei siti web mirati all'UE, ma l'aumento di Google è quattro volte maggiore.

Tabella 4.5: Cambiamenti nelle quote di mercato, per sottomercato, Google e Facebook contro tutte le altre aziende.

	(1) Advertising	(2) Social Media
EU	-4.2e-12 (.000061)	-2.7e-12 (.000263)
Post	5.3e-06 (.000023)	-.00007 (.000089)
Post × Facebook	.103305*** (.00003)	4.42552*** (.00003)
Post × Google	-.441782*** (.00003)	
Post × EU	-.000028 (.000029)	.000174 (.000174)
Post × EU × Facebook	-.054528*** (.000042)	-10.9839*** (.000088)
Post × EU × Google	1.83314*** (.000042)	
Observations	4050688	4050688
\bar{R}^2	.99528	.994232
Mean DV	.00158	.00158
% Effect nonEU Facebook	1.34712	8.43895
% Effect EU Facebook	.674025	-9.78537
% Effect nonEU Google	-1.58394	
% Effect EU Google	5.40281	

Una delle possibili ragioni di questa diversa entità di incremento può anche essere ricondotta al calo di reputazione in materia di privacy di cui Facebook ha sofferto negli ultimi anni. Si noti, ad esempio, che Facebook ha perso il 9,8% della sua quota di mercato, con siti web che si rivolgono ad un pubblico UE nel suo mercato più forte prima del GDPR (quota di mercato del 59,5%), ossia il mercato dei pulsanti like/share/login, che consentono alla società di tracciare gli utenti attraverso i siti web [64]. Per il principio di responsabilità congiunta questo calo di reputazione potrebbe aver avuto delle conseguenze nella scelta dei siti web di usare i vari tools offerti da Facebook.

4.2.1.6 Conclusioni dello studio

Pur con i limiti dovuti ad un intervallo di osservazione del campione molto piccolo e alla mancanza di dati a grana fine, questo studio documenta in primo luogo una diminuzione, in seguito all'entrata in vigore del GDPR, dell'utilizzo di tecnologie web di terze parti da parte dei siti web nel breve periodo, mentre nel lungo non si osserva un cambiamento nella tendenza all'aumento.

In secondo luogo, giunge alla conclusione secondo la quale, con l'introduzione del regolamento, i mercati delle tecnologie web sono diventati più concentrati, con l'azienda dominante - Google - che ha aumentato la sua quota di mercato. Tutte le altre aziende o non hanno visto alcun cambiamento nella quota di mercato o hanno subito delle perdite. Per gli utenti questo significa che mentre il numero di terze parti che chiedono l'accesso ai loro dati sta diminuendo, pochi stanno ricevendo più dati.

Diversi possono essere i motivi che spiegano queste conseguenze e lo stato favorevole di Google dopo il GDPR.

Innanzitutto, il fatto che il regolamento imponga come condizione per il trattamento dei dati personali il consenso dell'utente avvantaggia le imprese più grandi che offrono una più ampia gamma di servizi e sono quindi facilitati nel raccogliere il consenso degli utenti. Nella tabella 4.3 infatti si è dimostrato che la quota di mercato post-GDPR di un fornitore di tecnologia web aumenta con un numero maggiore di domini (cioè di servizi) e con un numero maggiore di sotto-mercati (cioè di categorie di servizi). Google si distingue come l'azienda con di gran lunga il maggior numero di domini, e di gran lunga il maggior numero di categorie di servizi [64].

Inoltre, è anche vero che, rispetto alle piccole aziende, Google e altre grandi aziende hanno potuto contare su molte più risorse da destinare alla conformità.

Infine, si può anche supporre che i proprietari di siti web che cercano di ridurre al minimo la loro esposizione optano per "scelte più sicure", abbandonando provider più piccoli che potrebbero avere più difficoltà a dimostrare la conformità.

Ad ogni modo è chiaro che l'aumento della concentrazione e la crescente quota di mercato della società dominante rappresentano di sicuro delle conseguenze indesiderate, che molto probabilmente il legislatore europeo non aveva in mente quando ha progettato il GDPR e suggeriscono al contempo che alcune delle sue più importanti implicazioni potrebbero non riguardare solo la privacy ma anche la concorrenza. Ciò solleva conseguentemente la questione di come la legge sulla privacy e la politica antitrust siano strettamente collegate e di come sia quindi sempre più difficile concettualizzare queste due aree del diritto come due aree distinte della legge con obiettivi, rimedi e meccanismi di applicazione diversi. Emblematico del fatto che le autorità stiano esplorando le connessioni tra antitrust e diritto della privacy nei mercati delle piattaforme digitali è il caso di cui si discuterà successivamente in cui l'antitrust tedesca (Bundeskartellamt) ha recentemente tentato di impedire a Facebook di combinare i dati degli utenti provenienti da diverse fonti.

4.3 Il caso: Facebook nel mirino del Bundeskartellamt

4.3.1 La privacy è una questione di concorrenza?

La raccolta e il trattamento dei dati personali nell'ambiente digitale ha una natura trasversale in ambito politico, in quanto, come si osserva dalla rappresentazione stilizzata in figura 5.9, tocca diversi settori legislativi: la legislazione della concorrenza, la protezione dei dati e la tutela dei consumatori.

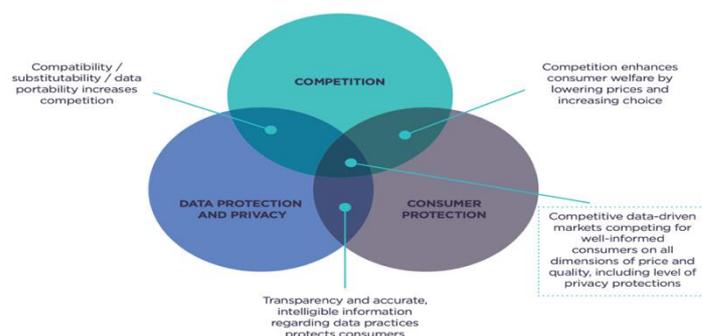


Figura 4.9: Sovrapposizione tra protezione dei dati, concorrenza e protezione dei consumatori [51]

In particolare, la discussione precedente ha dimostrato che esiste una notevole ***interdipendenza tra il diritto della concorrenza e il diritto sulla protezione dei dati e la privacy*** poiché quest'ultimo influisce sulla concorrenza e poiché l'eventuale potere di mercato influisce sia sulle scelte che gli interessati hanno realisticamente sia sui rischi per la privacy a cui sono esposti.

Riconoscendo, quindi, che qualsiasi questione relativa alla privacy e alla protezione dei dati personali non sia, in quanto tale, una questione che rientra nell'ambito di intervento delle autorità garanti della concorrenza ma debba essere inquadrata in modo più appropriato nel relativo quadro normativo, è impossibile non notare nel contesto dei mercati che si basano in modo profondo sulle informazioni personali dei consumatori una potenziale convergenza politica.

Di conseguenza, con l'aumento dei Big Data relativi ai consumatori due sono le domande che hanno suscitato un acceso dibattito:

1. La privacy è un problema di concorrenza?
2. E se sì, in che misura le considerazioni sulla privacy e le leggi sulla protezione dei dati devono filtrare nelle valutazioni del diritto della concorrenza?

La risposta alla prima domanda è abbastanza dibattuta, alcuni²⁹ pensano di no, ritenendo che il compito della politica della concorrenza sia esclusivamente quello di promuovere un mercato efficiente e il benessere e la scelta a disposizione dei consumatori e non quello di salvaguardare la protezione dei dati e la privacy dei consumatori; altri invece non sono d'accordo³⁰ e anzi invitano le autorità garanti della concorrenza a utilizzare il diritto della concorrenza per prevenire le violazioni delle leggi sulla protezione dei dati [66]. A riguardo vale la pena ricordare un crescente riconoscimento del fatto che i parametri non legati al prezzo che i consumatori apprezzano, e su cui quindi le imprese sono in competizione, possono essere altrettanto importanti per la concorrenza quanto il prezzo e debbano quindi essere considerati nelle valutazioni sulla concorrenza. È stato già discusso il tema secondo il quale, soprattutto nei mercati dei prodotti a prezzo zero in cui i consumatori "pagano" con i loro dati, la tutela della privacy possa essere vista come una dimensione della qualità.

Nel 2016 un report congiunto dell'autorità tedesca e francese [67] per la concorrenza, esaminando l'interazione tra il diritto della concorrenza e i dati affermava che "*... le politiche sulla privacy potrebbero essere considerate dal punto di vista della concorrenza ogniqualvolta tali politiche siano suscettibili di incidere sulla concorrenza, in particolare quando sono attuate da un'impresa dominante per la quale i dati servono come input principale dei suoi prodotti o servizi*".

La privacy come aspetto della concorrenza è stata discussa anche nel "Furman report" [68] del Regno Unito del 2019 che ha osservato che "*l'uso improprio dei dati dei consumatori e il danno alla privacy è probabilmente un indicatore di bassa qualità causato da una mancanza di concorrenza*".

²⁹ Grunes A.P., "Another Look at Privacy" (2013) 1111-1114, 1127; Sokol D.D. e Comerford R., "Does Antitrust Have A Role to Play in Regulating Big Data?".

³⁰ Autorité de la concurrence and Bundeskartellamt, "Competition Law and Data" (10 May 2016) 23

Inoltre, è interessante notare come negli ultimi anni in una serie di casi, soprattutto inerenti al controllo delle fusioni tra aziende che disponevano di una quantità enorme di dati e che avrebbero quindi potenzialmente potuto combinare le rispettive banche dati, l'aspetto della privacy sia stato preso in considerazione.

La possibilità che una concentrazione possa essere considerata anticoncorrenziale perché può portare ad una sostanziale diminuzione della concorrenza sulla privacy, o più in generale possa avere effetti negativi sul benessere dei consumatori a causa di una restrizione del livello di privacy nel mercato, è stata esplorata in alcune recenti decisioni di concentrazione nell'UE, ad esempio nel caso Google e Doubleclick³¹ nel 2008, nel caso Facebook-Whatsapp³² nel 2014 così come nel caso Microsoft-LinkedIn³³ nel 2016.

Sebbene in nessuno dei casi citati le questioni relative ai diritti dei dati dei consumatori siano state determinanti e non abbiano svolto un ruolo decisivo nelle decisioni della Commissione (nessuna delle concentrazioni è stata infatti bloccata sulla base di tali considerazioni), sembra esservi una crescente accettazione del fatto che la privacy possa essere rilevante per la valutazione della concorrenza [55].

L'interesse da parte delle autorità della concorrenza per le teorie del danno legate alla privacy potrebbe concentrarsi oltre che nell'ambito dell'applicazione ex ante, come nel caso del controllo delle fusioni, anche in quello dell'applicazione ex post, ad esempio nei casi in cui un'impresa dominante sfrutti il proprio potere di mercato fissando delle condizioni di sfruttamento eccessive non in relazione ai prezzi ma ad esempio in relazione ad un'eccessiva raccolta dei dati degli utenti. A tal proposito degno di nota è il caso dell'autorità della concorrenza tedesca (Bundeskartellamt) contro Facebook [69] il quale rappresenta un caso unico nel panorama della connessione tra la legislazione sulla protezione dei dati e la legislazione sulla concorrenza ed affronta il modo in cui il diritto della concorrenza potrebbe essere coinvolto in un'applicazione ex post in comportamenti di sfruttamento che portano a un danno alla privacy.

4.3.2 L'indagine del Bundeskartellamt

In quel che sembra essere la prima volta tra le autorità garanti della concorrenza in Europa di un caso di sfruttamento che riguarda i dati dei consumatori, nel marzo del 2016 l'autorità garante della concorrenza tedesca, il Bundeskartellamt, ha avviato un'indagine contro Facebook in relazione alle sue pratiche in materia di dati con il sospetto che la società avesse abusato del suo potere di mercato violando le norme sulla privacy e sulla protezione dei dati.

³¹ European Commission (2008), *Case No COMP/M.4731 – Google/ DoubleClick*, https://ec.europa.eu/competition/mergers/cases/decisions/m4731_20080311_20682_en.pdf.

³² Commission, F. (ed.) (2014), *FTC Notifies Facebook, WhatsApp of Privacy Obligations in Light of Proposed Acquisition*, <https://www.ftc.gov/news-events/press-releases/2014/04/ftc-notifies-facebook-whatsapp-privacy-obligations-light-proposed>.

³³ EC (2016), *Mergers: Commission approves acquisition of LinkedIn by Microsoft, subject to conditions*, https://ec.europa.eu/commission/presscorner/detail/en/IP_16_4284.

In particolare, l'obiettivo dell'indagine era quello di esaminare se Facebook avesse abusato della sua posizione dominante ed imposto ai suoi utenti condizioni sleali subordinando l'utilizzo del social network alla condizione che gli fosse consentito di accumulare illimitatamente ogni tipo di dati generati dall'utente durante l'utilizzo di siti web e applicazioni di terzi e di fonderli con l'account Facebook dell'utente stesso [69].

4.3.2.1 Analisi del contesto

Facebook Inc. è una società statunitense, fondata nel 2004 che sviluppa e gestisce diversi prodotti digitali, servizi online e applicazioni per smartphone.

Il suo prodotto principale è il social network *Facebook.com* che rappresenta con oltre 2,5 miliardi di utenti attivi al mese il leader indiscusso del settore a livello mondiale³⁴, come si evince anche dalla figura 4.10 seguente.



Figura 4.10: *Mapa mondiale dei social networks.* Fonte: vincos.it

Oltre a Facebook.com, Facebook offre anche Instagram, un servizio per la condivisione di foto e brevi video clip, il servizio di messaggistica WhatsApp, nonché "Facebook Business Tools" ossia una selezione di strumenti e prodotti gratuiti per operatori di siti web, sviluppatori, inserzionisti e altre aziende, da integrare nei propri siti web o applicazioni e offerte online tramite interfacce di programmazione (Application Programming Interfaces, API) predefinite da Facebook. La selezione comprende plugin sociali (pulsanti "Mi piace" o "Condividi"), login Facebook e altri servizi di analisi (Facebook Analytics) [70].

Società tra le più grandi al mondo, con una capitalizzazione di mercato di 583,7 miliardi di dollari US³⁵, fonda il suo modello di business sulla pubblicità online, settore da cui genera sostanzialmente tutti i suoi ricavi³⁶ dalla pubblicità online e nel quale insieme a Google si conferma leader indiscusso.

Proprio per questo motivo per la società i dati degli utenti costituiscono una risorsa essenziale per la personalizzazione degli annunci in base ai loro interessi, al loro potere di

³⁴ Nello specifico nel 3° trimestre del 2020 gli utenti attivi al mese sul social erano 2,7 miliardi con un incremento rispetto allo stesso periodo dell'anno precedente del 9,68%.

https://s21.q4cdn.com/399680738/files/doc_financials/2020/q3/FB-Q3-2020-Earnings-Presentation.pdf

³⁵ Fonte: www.statista.com : Biggest companies in the world by market capitalization 2020

³⁶ Nel 2019 dei 70,697 miliardi di dollari US, 69,655 miliardi di dollari erano rappresentati dalla pubblicità, con un aumento tra l'altro del 27% rispetto l'anno precedente. Fonte: <https://investor.fb.com/financials/default.aspx>

acquisto e alle loro condizioni di vita e sono da questa raccolti sia attraverso i propri servizi sia attraverso siti web e app di terzi.

La raccolta dati di Facebook

Proprio per via di questa serie di servizi offerti, Facebook è in grado di raccogliere un'ampia quantità di dati utente sia tramite le sue piattaforme che al di fuori di esse - si parla in quest'ultimo caso di *raccolta dati off-platform* - e di creare così profili utenti altamente dettagliati.

Nello specifico, i dati raccolti direttamente tramite i propri servizi includono [51]:

- dati di registrazione dell'utente: nome, data di nascita, sesso, e-mail, numero di telefono;
- dati sulla piattaforma Facebook: pagine con mi piace/seguite, rete di amici, contenuti e annunci con mi piace/commentato/cliccato, tempo trascorso a guardare gli annunci, contatti telefonici;
- dati dalle altre piattaforme possedute e gestite: da Instagram (pagine seguite, rete di amici, contenuti e annunci piaciuti/commentati/cliccati, persone messaggiate), da WhatsApp (persone messaggiate, numeri di telefono, contatti) e da Messenger (persone messaggiate, numeri di telefono, contatti, annunci cliccati).

Al contrario, i dati personali degli utenti off-platform sono quei dati raccolti mentre l'utente non utilizza i servizi di Facebook ma naviga su siti web e app in cui sono incorporati Business Tools di Facebook come i pulsanti "mi piace", "condividi", "login" nonché servizi analitici come "Facebook Analytics" per monitorare l'utilizzo e tracciare l'andamento delle campagne pubblicitarie.

Raccolta dati off-platform

Nella raccolta dati su siti web terzi Facebook non ha eguali se non da Google. Infatti, un documento pubblicato dalla Princeton University che ha analizzato il tracciamento online dei primi un milione di siti web ha rilevato che oltre il 70% dei siti web aveva un tracker di Google e oltre il 20% di Facebook (cfr. figura 4.11) [71].

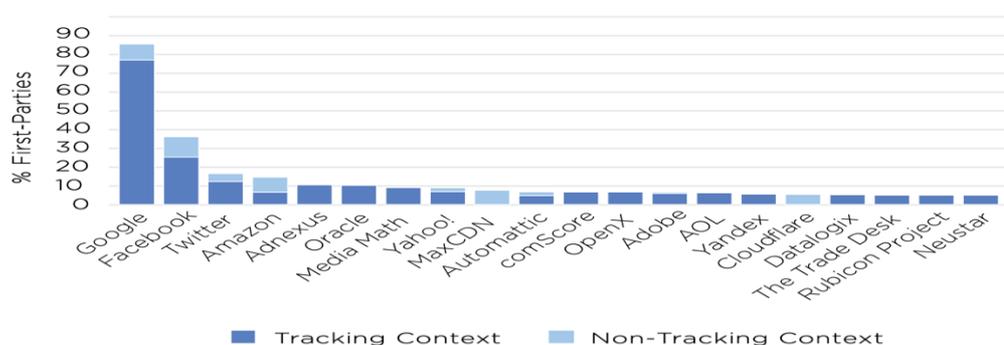


Figura 4.11: Online tracking by Google e Facebook. [51]

Lo studio ha rilevato che anche se un gran numero di tracker (81 000 tracker di terze parti sono stati individuati su almeno due prime parti) è presente nei primi un milione di siti web, Google, in particolare, e anche Facebook hanno tracker su un numero significativamente

maggiore di siti web rispetto ad altre aziende che utilizzano tracker di terze parti per raccogliere dati. In altre parole, Google e Facebook raccolgono molti più dati da siti web di terzi rispetto ad altre aziende [51].

Sebbene quindi già dall'utilizzo del proprio servizio Facebook.com raccolga molte informazioni sugli utenti, il flusso di dati aggiuntivi derivanti dalla combinazione dei dati di Instagram, WhatsApp e fonti terze è notevole e rappresenta un valore aggiuntivo. A maggior ragione se si considera appunto la sua ampia presenza sui siti web terzi e le dimensioni delle altre due piattaforme di sua proprietà, WhatsApp con oltre un miliardo di utenti in tutto il mondo e Instagram con oltre 500 milioni di utenti giornalieri in tutto il mondo [72]. Di conseguenza questi dati di alta qualità, dovuta alla granularità e alla specificità dei diversi attributi, se fusi in unico account permettono di ottenere un profilo dell'utente altamente dettagliato

4.3.2.2 La raccolta dati off-platform sotto la lente del Bundeskartellamt

Il caso in questione non riguarda i dati che Facebook raccoglie all'interno del suo social network quanto piuttosto i dati che questi raccoglie da fonti di terze parti, compresi i servizi digitali di sua proprietà (WhatsApp o Instagram) e i siti web e le applicazioni di terze parti che incorporano i suoi "business tools". In particolare, questi dati vengono trasmessi a Facebook.com tramite API nel momento in cui un utente richiama il sito web in cui questi tools sono incorporati, anche se l'utente non ha interagito con essi (ad esempio non ha premuto sul pulsante "mi piace") e anche se l'utente ha bloccato il web tracking. Dopodiché, Facebook.com procede a combinare questi dati e ad allegarli all'account social di un utente [72].

La scelta del Bundeskartellamt di concentrarsi sulla raccolta dati off-platform ha quindi una ragione, in quanto mentre gli utenti di Facebook.com possono aspettarsi che un certo livello di dati venga raccolto attraverso l'uso dei principali servizi del social network come contropartita per l'utilizzo del servizio che si finanzia tramite la pubblicità, al contrario, invece, gli utenti possono non essere consapevoli che Facebook raccolga dati sulle loro interazioni su altri siti e applicazioni apparentemente non correlate, e che utilizzi tali dati per creare profili utente e vendere agli inserzionisti l'inventario degli annunci.

Nel primo caso infatti l'autorità tedesca ha tenuto conto del fatto che un social network finanziato dalla pubblicità abbia come componente essenziale il trattamento di dati personali in una certa misura e non ha ravvisato alcun motivo per intervenire sulla base del divieto di pratiche abusive ai sensi del diritto della concorrenza, cosa che invece ha ritenuto di fare nel secondo caso [70].

Facebook.com, infatti, non solo effettua questa attività ma addirittura subordina l'utilizzo del social network all'accettazione da parte dell'utente, al momento della registrazione, di condizioni di servizio che prevedono tra le altre cose anche la raccolta dei dati degli utenti e dei loro dispositivi al di fuori delle attività di Facebook e l'assegnazione all'account social. Secondo il Bundeskartellamt il fatto che Facebook imponga una tale condizione sulla base di un principio "take-it-or-leave", ossia dia come unica scelta quella o di accettare l'"intero

pacchetto Facebook” o di fare a meno del servizio [73], costituisce una condotta pluri-offensiva, ossia lesiva al contempo del corretto funzionamento del mercato e delle disposizioni in materia di tutela della privacy e dei dati personali, in virtù soprattutto della sua posizione dominante nel settore dei social network. Pertanto, l’antitrust tedesca ha ritenuto opportuno istruire un procedimento contro la società in cui gli si contesta di aver abusato del suo potere di mercato imponendo ai suoi utenti delle condizioni che al contempo violerebbero non solo le norme a tutela dei dati personali, ma anche la lett. a) dell’art. 102 del TFUE in cui appunto si vieta l’abuso di posizione dominante.

4.3.2.3. Definizione del mercato rilevante

L’analisi giuridica del Bunderskattelamt parte con la prima fase tipica di un qualsiasi caso antitrust, ossia con la definizione del mercato rilevante. Ciò permette alle autorità di regolamentazione della concorrenza di individuare gli operatori del mercato, ossia fornitori, clienti e consumatori, di calcolare la dimensione totale dello stesso e la quota di mercato di ciascun fornitore con riferimento al prodotto o servizio rilevante nell’area interessata [74]. Questa attività considera in generale tre variabili:

1. il mercato del prodotto, compresi i prodotti e i servizi considerati intercambiabili o sostituibili dai consumatori;
2. il mercato geografico;
3. un orizzonte temporale, che riflette i cambiamenti nelle abitudini dei consumatori e gli sviluppi tecnologici.

A tal scopo, il Bunderskattelamt ha innanzitutto esaminato il modello di business di Facebook.com che corrisponde ad una piattaforma multi-versante finanziata attraverso la pubblicità mirata. I lati del mercato, come si osserva nella figura 4.12 seguente, sono diversi: da un lato vi sono gli utenti privati che utilizzano Facebook.com senza pagare un prezzo monetario, dall’altro, gli inserzionisti che gestiscono pubblicità mirate, ulteriori lati corrispondono poi agli editori che utilizzano Facebook.com per promuovere la propria attività con le proprie pagine Facebook sulle quali pubblicano contenuti editoriali e si collegano con gli utenti e gli sviluppatori i quali possono integrare i prodotti Facebook nei loro siti web o nelle loro applicazioni. Tra le parti esistono effetti di rete indiretti [70].

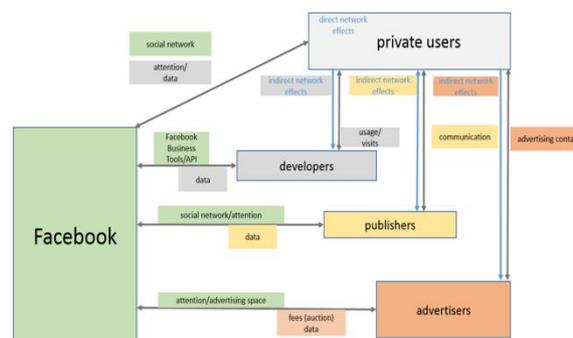


Figura 4.12: Lati della piattaforma multi-versante di Facebook. [70]

Nei mercati digitali nel determinare la sostituibilità l'autorità ha generalmente una certa flessibilità in quanto considera non solo le caratteristiche del prodotto e la sua destinazione d'uso, ma anche altri fattori, tra cui il punto di vista dei clienti e dei concorrenti, le prove delle preferenze dei clienti e l'esistenza di diverse categorie di clienti per il prodotto [74].

Il Bundeskartellamt, in particolare, sulla base del concetto di sostituibilità dal lato della domanda, ha definito come mercato rilevante il **mercato dei social network con utenti privati** come lato opposto del mercato rilevante. Ha ritenuto, infatti che, poiché nessuno dei suddetti gruppi di utenti di Facebook ha esigenze simili a quelle del gruppo di utenti privati, essi devono essere attribuiti ad altri mercati. Inoltre, poiché dalle indagini condotte l'autorità ha osservato che gli utenti tedeschi utilizzano prevalentemente i social network per rimanere in contatto con amici e conoscenti in Germania, il **mercato geografico rilevante** considerato è quello **nazionale** [70].

Definito il mercato rilevante, l'autorità ha proceduto ad individuare i potenziali concorrenti, nel farlo ha considerato l'elevato grado di differenziazione dei prodotti dei social media e le varie sovrapposizioni delle loro funzionalità, nonché i forti effetti diretti di rete, ed è giunta alla conclusione che, poiché Google+ è ormai scomparso dal mercato, l'insieme dei concorrenti di Facebook comprendeva solo alcuni piccoli fornitori tedeschi di social network i cui prodotti erano in generale paragonabili a quelli forniti da Facebook.

Altri servizi come LinkedIn, Youtube, Snapchat, Twitter, Pinterest e Instagram non sono stati inclusi nel mercato del prodotto rilevante in quanto o non sono in diretta concorrenza con Facebook o perché, seppur lo sono in parte o in relazione a singole funzioni, rispondono principalmente a esigenze diverse dal punto di vista degli utenti privati (ad esempio LinkedIn è stato concepito per soddisfare le esigenze professionali degli utenti) [72].

Infine, l'autorità ha adottato anche una visione dinamica nel valutare oltre la concorrenza attuale anche quella potenziale, ed è giunta alla conclusione che in servizi caratterizzati da forti effetti di rete sia diretti che indiretti, come nel caso in questione, non è sufficiente disporre di una "massa critica" di utenti o di competenze tecniche, finanziarie e personali per poter entrare nei mercati vicini ed avere successo come sul mercato originario e Google+ ne è una dimostrazione [70].

4.3.2.4 Analisi del potere di mercato

Dopo aver definito il mercato rilevante la fase successiva è la valutazione del potere di mercato. Solitamente il fattore determinante nella valutazione della quota di mercato di un'impresa è il fatturato della stessa, o il volume o il valore delle vendite totali del prodotto rilevante nell'area interessata. Tuttavia, in un mercato digitale una quota di mercato detenuta da un fornitore di un servizio online gratuito non può essere facilmente calcolata con riferimento ai dati sulle vendite o ai suoi volumi tradizionali. Per cui per misurare la quota di mercato rilevante in un settore specifico per l'economia digitale, le linee guida della Commissione Europea raccomandano di scegliere il criterio più appropriato alla luce delle caratteristiche del mercato [74].

A tal scopo in primo luogo il Bundeskartellamt ha esaminato la quota di mercato di Facebook basata sull'utente sul mercato rilevante che è risultata molto alta. Il numero di

utenti attivi (giornalieri o mensili) sono stati infatti considerati come un indicatore chiave e una misura rilevante per valutare l'importanza concorrenziale della rete e il successo di mercato di un social network che si ritiene essere dipendente dall'intensità dell'utilizzo. Nello specifico l'autorità ha rilevato che, con 23 milioni di utenti attivi giornalieri e 32 milioni di utenti attivi mensili nel 2018, Facebook aveva una quota di mercato di oltre il 95% tra gli utenti attivi giornalieri e di oltre l'80% tra gli utenti attivi mensili [72].

Oltre alla sua elevata quota di mercato, il Bundeskartellamt considera un'altra serie di criteri³⁷ che sono significativi per il caso Facebook e che forniscono la prova del suo potere di mercato. Elementi chiave del test di dominanza del mercato sono infatti innanzitutto i forti effetti di rete diretti che caratterizzano il modello di business di Facebook e che portano ad un effetto di lock-in che rende difficile per gli utenti il passaggio ad un altro social network che non sia altrettanto attrattivo.

Lo stesso si può dire per gli effetti di rete indiretti tra utenti ed inserzionisti che aumentano le barriere all'ingresso sul mercato rendendo così difficile l'ingresso di altre piattaforme finanziate dalla pubblicità. Per potere infatti entrare nel mercato con successo con un prodotto ad-supported un nuovo concorrente dovrebbe raggiungere un numero critico di utenti privati perché altrimenti il prodotto non sarebbe sufficientemente attraente per il lato pubblicitario. Tuttavia, gli effetti diretti di rete rendono appunto difficile raggiungere un numero così critico di utenti.

Altri fattori di potere di mercato sono rappresentati dalle elevate economie di scala di cui Facebook gode, da un grado trascurabile di utilizzo parallelo di diversi servizi diversi ("multi-homing"), e da un accesso superiore rispetto ai concorrenti ai dati personali degli utenti.

Tra l'altro il Bundeskartellamt ha osservato come il mercato tedesco dei social network fosse caratterizzato da una tendenza al ribasso delle quote di mercato basate sugli utenti dei concorrenti, alcuni dei quali sono usciti completamente dal settore, mentre le cifre di Facebook continuavano a salire [70].

In definitiva, sulla base di una valutazione complessiva di tutti questi fattori, l'autorità tedesca ha concluso che Facebook era l'impresa dominante sul mercato nazionale dei social network per utenti privati, in quanto dotata di un campo d'azione in quel mercato che non era sufficientemente controllato dalla concorrenza.

4.3.3 La condotta di Facebook è in violazione della legge?

L'autorità garante della concorrenza tedesca nel caso Facebook, nel determinare l'esistenza di un abuso di posizione dominante, sembra procedere ad un ragionamento in due fasi.

4.3.3.1 Esame della compatibilità delle pratiche di Facebook con il GDPR

³⁷ Con l'ultima modifica della legge tedesca sulla concorrenza (il Gesetz gegen Wettbewerbsbeschränkungen o GWB) nel 2017 il legislatore ha precisato i metodi di valutazione del potere di mercato considerando, in particolare, come suoi fattori rilevanti l'accesso a dati rilevanti dal punto di vista della concorrenza, le economie di scala basate sugli effetti di rete, il comportamento degli utenti che possono utilizzare più servizi diversi o un solo servizio e il potere della pressione concorrenziale determinata dall'innovazione [70].

Innanzitutto, l'autorità tedesca parte con un'analisi piuttosto approfondita dei termini, delle condizioni di servizio e delle politiche sui dati di Facebook.com e verifica la loro ammissibilità alla luce dei principi stabiliti dagli articoli 6 e 9 del GDPR.

Come è stato già discusso nella sezione 4.1.3.2, il GDPR stabilisce che i dati personali possono essere trattati solo per scopi specifici e sulla base del consenso della persona interessata, consenso che per essere ritenuto valido deve essere "liberamente dato". In particolare, il GDPR approfondisce ulteriormente questo concetto affermando in uno dei suoi considerando che il consenso necessario per il trattamento dei dati personali "si presume che non venga dato liberamente se... l'esecuzione di un contratto, compresa la fornitura di un servizio, è subordinata al consenso, nonostante tale consenso non sia necessario per tale esecuzione." (considerando 43 del GDPR).

In particolare, in considerazione della natura "take-it-or-leave" che esprime il carattere condizionale dello scambio e dell'evidente squilibrio, in virtù soprattutto della posizione dominante di Facebook, tra il titolare del trattamento dei dati e il soggetto interessato, l'autorità ha ritenuto che il consenso non potesse essere considerato come "liberamente dato". Il consenso degli utenti sarebbe cioè effettivo solo se la fornitura del servizio di Facebook.com non fosse subordinata ad esso.

L'autorità ha inoltre respinto la motivazione di Facebook, secondo la quale una tale elaborazione dei dati sarebbe necessaria per fornire il servizio e per soddisfare i suoi legittimi interessi.

A tal proposito infatti il Bunderskattelamt ha sostenuto che in realtà l'elaborazione di dati raccolti da altri servizi di sua proprietà o di terzi, nella misura determinata da Facebook nelle sue condizioni generali di contratto, non fosse necessaria né per l'esecuzione del contratto con l'utente né per monetizzare la rete attraverso la pubblicità personalizzata, in quanto Facebook avrebbe potuto raggiungere un alto grado di personalizzazione basandosi in larga misura sui dati degli utenti elaborati nell'ambito della gestione del social network stesso.

Di particolare interesse nell'esaminare la compatibilità delle pratiche di Facebook con il GDPR è anche il fatto che l'autorità tedesca per la concorrenza abbia costruito la sua teoria del danno sulla base dei diritti dei consumatori all'autodeterminazione informativa e alla privacy, in quanto questi sono proclamati e protetti dalla legge sulla protezione dei dati e non si sia basata, come solitamente fanno le autorità di concorrenza, solo sulle preferenze dei consumatori in quanto queste si rivelano sul mercato.

In conclusione, ricapitolando, il Bunderskattelamt ha ritenuto che l'esistenza di una posizione dominante fosse rilevante per valutare la validità del consenso, che quindi le giustificazioni basate sul GDPR per la condotta di Facebook non fossero applicabili, in considerazione del "grave squilibrio" tra gli interessi di Facebook e la tutela dei diritti fondamentali degli utenti, e che pertanto esisteva nelle pratiche di trattamento dati off-platform di Facebook.com una violazione dei principi del GDPR. Ciò potrebbe tuttavia non aver comportato in modo automatico una violazione del diritto della concorrenza, per l'autorità procede durante la seconda fase dell'analisi alla valutazione di questa condotta sotto il profilo della concorrenza [75].

4.3.3.2 Esame della condotta sotto il profilo della concorrenza

In secondo luogo, il Bunderskattelamt ha analizzato anche come la violazione della protezione dei dati potesse costituire una violazione ai sensi delle disposizioni del diritto della concorrenza, in particolare dell'articolo 19, paragrafo 1 del GWB tedesco che rappresenta in sostanza l'equivalente nel diritto comunitario dell'art. 102 del TFUE. Nello specifico quest'ultimo vieta l'abuso di posizione dominante il quale può esplicitarsi o in un abuso di sfruttamento nei rapporti che coinvolgono i consumatori o i fornitori o in un abuso di esclusione relativamente invece ai concorrenti.

Per quanto riguarda gli abusi di sfruttamento, la legislazione antitrust europea non solo "sorveglia" il campo del comportamento delle imprese dominanti in materia di prezzi, ma anche quello di qualsiasi altra pratica di sfruttamento che possa eventualmente derivare dal loro maggiore potere economico, a maggior ragione nel caso in cui il compenso per il servizio non sia di carattere monetario ma consista nella raccolta dei dati personali dell'utente.

In linea, quindi, con l'idea secondo la quale non solo i prezzi eccessivi, ma anche le condizioni contrattuali inappropriate costituiscono un abuso di sfruttamento, l'autorità tedesca ha ritenuto che la politica dei dati di Facebook, che gli consente di raccogliere dati relativi agli utenti da fonti esterne a Facebook e di fonderli con i dati raccolti sul social network, costituisce un abuso di posizione dominante sotto forma di condizioni commerciali di sfruttamento ai sensi dell'articolo 19, paragrafo 1, GWB.

A riguardo il Bunderskattelamt non ha eseguito - come spesso accade per l'analisi di altre pratiche di sfruttamento, come ad esempio i prezzi eccessivi - alcuna quantificazione economica della condotta abusiva in termini di confronto tra il danno e il beneficio per il consumatore netto e un controfattuale, ritenendo che tale quantificazione economica fosse "quasi impossibile" in questo caso [69]. Tuttavia, ha sostenuto che ciò non può mettere in discussione la constatazione di un danno per i consumatori. L'autorità ha ritenuto, infatti, che tale danno risieda innanzitutto nella perdita di controllo da parte dei consumatori dei loro dati personali in quanto gli utenti non sono più in grado di percepire quali dati, da quali fonti e per quali scopi vengono combinati con i dati degli account Facebook. In secondo luogo il Bunderskattelamt osserva che *"si può presumere che il comportamento contestato possa anche portare a un potenziale danno per gli utenti in termini economici"*, ad esempio nel caso di prezzi personalizzati, in termini finanziari qualora Facebook divulghi i dati a terzi e questo porti ad un rischio di furto di identità, estorsione o frode, ed infine anche in termini di danni non materiali nella misura in cui la raccolta di dati riveli informazioni sensibili, quali malattie, opinioni politiche e l'orientamento sessuale.

Ovviamente, oltre al danno per i consumatori, queste pratiche comportano anche un danno per i concorrenti e presentano il rischio di rafforzare ulteriormente il potere di mercato di Facebook nella misura in cui grazie ad esse Facebook è riuscito ad ottimizzare il proprio servizio, a "legare" quindi più utenti sia privati che business alla propria rete per via degli effetti di rete e degli effetti di lock-in e a creare in tal modo delle barriere all'ingresso nel mercato per i nuovi operatori che non sono in grado di accumulare un tale ammontare di dati.

In definitiva per quanto riguarda l'esame della condotta di Facebook sotto il profilo della concorrenza, l'antitrust tedesca ha sostenuto che grazie all'elaborazione inappropriata dei dati e alla loro combinazione con gli account di Facebook questi ha acquisito un vantaggio competitivo rispetto ai concorrenti in modo illecito e ha aumentato le barriere all'ingresso sul mercato, il che a sua volta ha assicurato e rafforzato il suo potere di mercato nei confronti dei clienti finali.

4.3.4 Conclusione del caso e decisione dell'antitrust tedesca

L'indagine si è conclusa il 6 febbraio 2019 con la constatazione da parte del Bundeskartellamt dell'abuso di posizione dominante da parte di Facebook sul mercato dei social media per quanto riguarda la raccolta di dati off-platform e la conseguente imposizione a Facebook di modificare le sue pratiche di raccolta ed elaborazione dei dati entro 12 mesi.

Nello specifico l'autorità ha vietato a Facebook la pratica di subordinare l'uso del social network da parte degli utenti privati residenti in Germania alla condizione che Facebook possa raccogliere i dati relativi agli utenti da altri servizi aziendali o esterni e combinarli con gli account utente di Facebook.com. Solo nel caso in cui l'utente acconsente, Facebook può ancora combinare i dati nel modo completo sopra descritto, altrimenti, senza il consenso dell'utente, l'elaborazione dei dati deve avvenire di norma in un processo separato internamente [72].

La società ha presentato ricorso contro questa decisione presso il tribunale regionale superiore di Düsseldorf, che nell'agosto 2019 ha sospeso l'ordinanza in quanto non ha accettato l'approccio adottato dal Bundeskartellamt durante il procedimento.

In particolare, non ha accettato innanzitutto che una possibile violazione delle norme sulla privacy possa far scattare automaticamente una violazione delle norme antitrust nel caso di una società dominante. In secondo luogo, il Bundeskartellamt è stato anche criticato per non aver fornito alcuna analisi controfattuale e per non aver dimostrato né che i consumatori apprezzano la privacy nel mercato dei social network in Germania né come le pratiche dei dati di Facebook hanno danneggiato la concorrenza [75].

La sospensione dell'ordinanza ha esonerato quindi Facebook dall'attuare la decisione del Bundeskartellamt. Tuttavia, quest'ultima in seguito a questo evento ha presentato ricorso contro la sospensione alla Corte di giustizia federale, la quale a giugno del 2020 si è espressa a favore dell'antitrust tedesco, sferrando così un duro colpo al colosso di Menlo Park³⁸.

In conclusione, nel panorama dei Big Data e dell'interdipendenza tra concorrenza e privacy questo caso ha sicuramente una rilevanza unica in quanto la sua decisione è stata la prima che ha sostenuto una teoria del danno relativa alla riduzione della privacy come abuso di posizione dominante [55].

³⁸https://www.repubblica.it/esteri/2020/06/24/news/germania_sentenza_facebook_vince_la_tutela_dei_consumatori-260055587/

In particolare, a tal proposito il Bunderskattelamnt tedesco ha seguito un approccio secondo il quale una violazione di disposizioni di legge diverse da quelle relative alla concorrenza - nello specifico del caso in questione relative alla privacy e alla protezione dei dati - diventa un problema di diritto della concorrenza se la violazione è il risultato di una posizione dominante o conferisce un vantaggio competitivo che consente all'impresa dominante di falsare la concorrenza [75].

L'autorità, pur nelle difficoltà di casi come questi, dovuti ad esempio all'assenza di un prezzo monetario nullo così come a fattori legati alla domanda quali i bias comportamentali, si è adoperata per proporre un'interpretazione coerente del diritto della concorrenza al diritto della protezione dei dati e per stabilire un quadro concettuale che consentisse l'applicazione simultanea di entrambi i settori del diritto.

Perciò questo caso rappresenta un precedente per cui non si può escludere che la sua decisione possa avere un'influenza sulle condizioni preliminari per la raccolta e l'aggregazione dei dati da parte delle aziende dominanti in futuro. La richiesta ai consumatori di fornire il consenso al trattamento dei dati da parte di servizi specifici di un'impresa dominante ma di non subordinare ad esso l'utilizzo di un servizio potrebbe, infatti, aiutare a controbilanciare l'auto-rafforzamento della loro posizione dominante attraverso l'accesso preferenziale ai dati [40].

Conclusioni

Il presente lavoro di tesi ha innanzitutto esaminato come nell'era digitale odierna il trattamento dei dati personali sia un'attività critica per la competitività tra le imprese digitali in quanto consente di utilizzare le informazioni sugli utenti per adattare i servizi alle loro specifiche esigenze e per monetizzare gli stessi in modo più efficiente. In particolare, in presenza di determinate condizioni, si osserva una tendenza di spostamento dei mercati digitali verso una dinamica del tipo "winner-take-all", dove le aziende che raggiungono la leadership di mercato, o riescono a controllarne una quota significativa, possono ribaltare il mercato a loro favore, diventando molto difficili da "detroneizzare" e creando una serie di problemi nell'area della concorrenza.

La capacità di raccogliere ed elaborare enormi moli di dati dai consumatori e di offrire loro servizi basati su essi è, infatti, uno degli elementi alla base del successo dirompente del poker da oltre cinquemila miliardi di dollari di capitalizzazione formato da Google, Amazon, Facebook e Apple. L'emergere di questi *data-opoli*, ossia di questo gruppo di società che controllano attraverso le loro piattaforme un volume significativo e una varietà di flussi di dati personali incomparabile per altre imprese, ha posto l'attenzione però non solo sui problemi di concorrenza che ne derivano ma anche relativamente ai rischi in termini di privacy che il trattamento di tali dati e la concentrazione in mano di pochi attori potrebbero comportare.

Ecco perché un'analisi che voglia essere il più possibile completa sul fenomeno dei Big Data non può prescindere dall'affrontare anche l'argomento relativo alla tutela della privacy. Nel caso di dati personali, le due aree del diritto sono infatti profondamente interdipendenti e si influenzano a vicenda.

A tal proposito, dall'analisi condotta si è potuto osservare innanzitutto come le norme e i regolamenti, nello specifico il GDPR europeo, che disciplinano la raccolta e l'accesso ai dati dei consumatori, potrebbero avere, in parallelo, un impatto anche sulla dimensione economica delle imprese e sulla concorrenza, di conseguenza alcune delle più importanti implicazioni del GDPR potrebbero non riguardare la privacy ma la politica antitrust. In particolare, le norme volte ad aumentare la tutela della privacy, potrebbero avere l'effetto perverso di favorire le big tech e di ridurre la concorrenza, laddove i costi di conformità a cui devono incorrere indistintamente tutte le imprese, siano esse grandi o piccole, siano elevati. Così come altre disposizioni, quali ad esempio il diritto alla portabilità dei dati, pensato appositamente per favorire la concorrenza e ridurre gli effetti di lock-ins degli utenti, potrebbe non sortire l'effetto desiderato, laddove i costi di attuazione siano elevati e il caso d'uso per i consumatori sia limitato.

A minare poi l'efficacia di questi strumenti e a rendere imprevedibili le risposte dei consumatori ai rimedi potrebbero concorrere altri fattori dal lato della domanda, tra cui i pregiudizi decisionali e il paradosso della privacy. Di conseguenza, le formalità del consenso, da sole, potrebbero fornire un meccanismo non ottimale per la tutela della privacy, a maggior ragione in mancanza di concorrenza.

In mercati caratterizzati da una scarsa concorrenza, infatti, probabilmente modelli di protezione dei dati basati sul consenso potrebbero non risultare particolarmente efficaci e anzi potrebbero dar vita al persistere della tendenza della maggiore concentrazione dei dati nelle mani dell'impresa con potere di mercato e del conseguente rafforzamento della sua posizione dominante.

L'analisi condotta ha dunque portato alla conclusione che, in un mondo in cui l'elaborazione dei dati personali e la capacità tramite questi di definire profili e prevedere i comportamenti degli utenti, costituisce un'attività essenziale dei mercati online altamente concentrati, è sempre più difficile concettualizzare la legge antitrust e la legge sulla privacy come due aree distinte con obiettivi, rimedi e meccanismi di applicazione diversi e senza implicazioni l'una per l'altra. Pertanto, in questo scenario, adottare isolatamente un'unica "lente politica" rischia di non affrontare il problema in modo efficace e in alcuni casi può persino portare ad effetti perversi indesiderati in altre aree. Riunire quindi le diverse prospettive politiche è essenziale per giungere a risultati ottimali sia in termini di benessere che di protezione dei consumatori. A tal proposito, l'indagine del Bundeskartellamt sulla raccolta dati off-platform di Facebook, costituisce un esempio di come le autorità della concorrenza stiano esplorando le connessioni tra antitrust e diritto della privacy nei mercati delle piattaforme digitali. Il caso in questione ha dimostrato come il diritto della concorrenza deve tenere conto della legislazione sulla protezione dei dati e l'interpretazione e l'attuazione della legge sulla protezione dei dati dovrebbe presumibilmente considerare sia l'esistenza o l'assenza di concorrenza sia gli effetti sulla concorrenza.

In conclusione si ritiene che solo una maggiore cooperazione e coordinamento tra le autorità preposte alla concorrenza, alla tutela dei consumatori e alla protezione dei dati, unita ad una maggiore comprensione delle dinamiche concorrenziali nei mercati che coinvolgono i dati dei consumatori, potrebbe aiutare nella progettazione di rimedi e nella verifica degli interventi prima della loro attuazione in modo tale da tenere in dovuta considerazione le possibili implicazioni su altre aree e da limitare le conseguenze indesiderate, nonché nell'individuazione e nello sviluppo di standard condivisi accettati per promuovere la privacy come un potenziale aspetto della concorrenza sulla qualità nella valutazione dei casi di applicazione della normativa sulla concorrenza nei mercati che riguardano i dati dei consumatori, soprattutto laddove questi offrano servizi a prezzo zero.

Bibliografia

- [1] Manyika, J., Chui, M., Brown, B., Bughin, J., Dobbs, R., Roxburgh, C. e Byers A. H. (2011) *Big Data: The Next Frontier for Innovation, Competition, and Productivity*, McKinsey Global Institute.
- [2] Gantz J. e Reinsel D. (2012), *The Digital Universe in 2020: Big Data, Bigger Digital Shadows, and Biggest Growth in the Far East*, IDC iView.
- [3] De Mauro, A., Greco, M. e Grimaldi, M. (2016), *A formal definition of Big Data based on its essential features*, Library Review, Vol. 65 No. 3, pp. 122-135.
- [4] Laney D. (2001), *3D Data Management: Controlling Data Volume, Velocity, and Variety*, META Group Report.
- [5] Schroeck M., Shockley R., Smart J., Romero-Morales D., Tufano P. (2012), *Analytics: the real-world use of big data. How innovative enterprises extract value from uncertain data*, Executive Report, IBM Institute for Business Value.
- [6] OECD (2015), *Data-Driven Innovation: Big Data for Growth and Well-Being*, OECD Publishing, Paris, <https://doi.org/10.1787/9789264229358-en>.
- [7] Reinsel D., Gantz J., Rydning J. (2017), *Data Age 2025: The evolution of Data to life-critical. Don't focus on Big Data; Focus on the Data that's Big*, White Paper IDC, Sponsored by Seagate.
- [8] NVP, NewVantage Partners (2016), *Big Data: An update on the adoption of Big data in the Fortune 1000*.
- [9] AGCOM (2018), *Big Data-Interim report nell'ambito dell'indagine conoscitiva di cui alla delibera No. 217/17/CONS*.
- [10] AGCOM, AGCM e GARANTE PER LA PROTEZIONE DEI DATI PERSONALI (2020), *Indagine conoscitiva sui Big Data*.
- [11] OFCOM (2019), *Online Nation*, Annual Report.
- [12] Abecassis D., Morgan R. (2019), *The use of data by online services*, Report for OFCOM, Analysys Mason.
- [13] Robertson V. (H.S.E) (2019), *Excessive data collection: Privacy considerations and abuse of dominance in the era of Big Data*. Working Paper
- [14] OECD (2020), *Consumer Data Rights and Competition – Background note by the Secretariat*.
- [15] Purra J. e Carlsson N. (2016), *Third-Party Tracking on the Web: A Swedish Perspective*, <http://dx.doi.org/10.1109/LCN.2016.14>.

- [16] Binns R., Lyngs U., Van Kleek M., Zhao J., Libert T., Shadbolt N. (2018), *Third Party Tracking in the Mobile Ecosystem*, <http://dx.doi.org/10.1145/3201064.3201089>.
- [17] Federal Trade Commission (2014), *Data Brokers: A Call for Transparency and Accountability*.
- [18] OECD (2016), *Big Data: Bringing competition policy to the digital era – Background note by the Secretariat*.
- [19] Autorité de la concurrence, Bundeskartellamt (2016), *Competition Law and Data*.
- [20] Acquisti A., Taylor C. e Wagman L., (2016), *The Economics of Privacy*, Journal of Economic Literature 2016, 54(2), 442–492 <http://dx.doi.org/10.1257/jel.54.2.442>
- [21] Abrams M. (2014), *The Origins of Personal Data and its Implications for Governance*, The Information Accountability Foundation.
- [22] OECD (2020), *Recommendation of the Council concerning Guidelines Governing the Protection of Privacy and Transborder Flows of Personal Data*, OECD/LEGAL/0188. <https://legalinstruments.oecd.org/public/doc/114/114.en.pdf>
- [23] Bourreau M., de Streel A., Graed I. (2017), *Big Data and Competition Policy: Market power, personalized pricing and advertising*, Project Report, Centre on Regulation in Europe (CERRE).
- [24] Duhigg C. (2012), *How Companies Learn Your Secrets*, The New York Times, URL: www.nytimes.com.
- [25] Johnson J, (2020), *Structured vs Unstructured Data in 2020: A Shift in Privacy*, BMC Machine Learning & Big Data Blog, URL: www.bmc.com
- [26] Szczepański M., (2020), *Is data the new oil? -Competition issues in the digital economy*, European Parliamentary Research Service.
- [27] Tucker C., (2019), *Digital Data, Platforms and the Usual [Antitrust] Suspects: Network Effects, Switching Costs, Essential Facility*, Review of Industrial Organization.
- [28] Lerner A. V., (2014), *The Role of “Big Data” in Online Platform Competition*. URL: <http://ssrn.com/abstract=2482780>
- [29] *Privacy and competitiveness in the age of big data. The interplay between data protection, competition law and consumer protection in the Digital Economy*. European Data Protection Supervisor, Marzo 2014. URL: http://europa.eu/rapid/press-release_EDPS-14-6_en.htm?locale=en
- [30] Stuck M. E., Grunes A. P., (2016), *Big Data and Competition Policy*.

- [31] Tucker D.S, Wellford H.B., (2014), *Big Mistakes Regarding Big Data*, Antitrust Source.
- [32] Rubinfeld D. L., Gal M. S. (2017), *Access Barriers to Big Data*, Arizona Law Review.
- [33] Chellappa R. K., Sin R.G. (2005), *Personalization versus Privacy: An Empirical Examination of the Online Consumer's Dilemma*, Article in Information Technology and Management, DOI: 10.1007/s10799-005-5879-y.
- [34] Benlian A., (2015), *Web Personalization Cues and Their Differential Effects on User Assessments of Website Value*, Article in Information Technology and Management, DOI: 10.1080/07421222.2015.1029394
- [35] Adomavicius G., Tuzhilin A. (2005), *Toward the Next Generation of Recommender Systems: A Survey of the State-of-the-Art and Possible Extensions*, IEEE Transactions On Knowledge And Data Engineering, Vol. 17, No. 6.
- [36] Adomavicius G., Bockstedt J., Curley S., Zhang J., (2017), *Effects of Online Recommendations on Consumers' Willingness to Pay*.
- [37] Hinz O., Eckert J., (2010), *The Impact of Search and Recommendation Systems on Sales in Electronic Commerce*. DOI 10.1007/s12599-010-0092-x
- [38] Boerman S.C., Kruikemeier S., Zuiderveen Borgesius F.J., (2017), *Online Behavioral Advertising: A Literature Review and Research Agenda*, Journal of Advertising, 46:3, 363-376, DOI: 10.1080/00913367.2017.1339368
- [39] Bleier A., Eisenbeiss M., (2015), *Personalized Online Advertising Effectiveness: The Interplay of What, When, and Where*, Marketing Science, DOI: 10.1287/mksc.2015.0930.
- [40] Crémer J., de Montjoye Y., Schweitzer H., (2019), *Competition policy for the digital era*.
- [41] Calvano E., Polo M., (2019), *Market power, competition and innovation in digital markets: A survey*, Information Economics and Policy, URL: <https://doi.org/10.1016/j.infoecopol.2020.100853>
- [42] Evans D.S, Schmalensee Richard., (2012), *The Antitrust Analysis of Multi-Sided Platform Businesses*, Coase-Sandor Institute for Law & Economics Working Paper No. 623.
- [43] Commissione Europea, (2015), *Public Consultation on the regulatory environment for platforms, online intermediaries, data and cloud computing and the collaborative economy*, pag.5, URL: <https://ec.europa.eu/digital-agenda/en/news/public-consultation-regulatory-environment-platforms-online-intermediaries-data-and-cloud>.

- [44] Commissione Europea, (2016), *Le piattaforme online e il mercato unico digitale. Opportunità e sfide per l'Europa*. URL: <https://eur-lex.europa.eu/legal-content/IT/TXT/PDF/?uri=CELEX:52016DC0288&from=EN>
- [45] Rochet J., Tirole, J., (2006), *Two-sided markets: a progress report*, RAND Journal of Economics.
- [46] Parker G.G, Van Alstyne M. W., (2005), *Two-Sided Network Effects: A Theory of Information Product Design*, Management Science, Vol. 51, No. 10, pp. 1494–1504.
- [47] Eisenmann T., Parker G., Van Alstyne M. W., (2006), *Strategies for two-sided markets*, Harvard Business Review.
- [48] Evans P.C, Gawer A., (2016), *The Rise of the Platform Enterprise. A Global Survey*, The Center for Global Enterprise.
- [49] Valletti T., Prat A., (2018), *Attention Oligopoly*, URL: <https://doi.org/10.2139/ssrn.3197930>
- [50] Krämer J., Schnurr D., (2019) *Competition Policy in Platform and Data-driven Markets: Long-term Efficiency & Exploitative Conducts. Contribution to the Call: "Shaping competition policy in the era of digitization"*.
- [51] Australian Competition & Consumer Commission (ACCC), (2019), *Digital Platforms Inquiry - Final Report*.
- [52] Kerber W., (2016), *Digital Markets, Data, and Privacy: Competition Law, Consumer Law, and Data Protection*, Joint Discussion Paper Series in Economics, No. 14-206, URL:https://www.uni-marburg.de/fb02/makro/forschung/magkspapers/index_html%28magks%29.
- [53] European Commission (2015) *Special Eurobarometer 431 - Data Protection Report*, URL:http://ec.europa.eu/commfrontoffice/publicopinion/archives/ebs/ebs_431_en.pdf
- [54] Acquisti A., Taylor C., Wagman L., (2016), *The Economics of Privacy*, Journal of Economic Literature 2016, 54(2), 442–492, <http://dx.doi.org/10.1257/jel.54.2.442>
- [55] OECD (2020), *Consumer Data Rights and Competition - Background note by the Secretariat*, Meeting of the Competition Committee on 12 June 2020. URL: <http://www.oecd.org/daf/competition/consumer-data-rights-and-competition.htm>.
- [56] Filistrucchi L., Geradin D., van Damme E., Affeldt P., (2014), *Market Definition in Two-Sided Markets: Theory and Practice*, Journal of Competition Law & Economics, Volume 10, Issue 2, Pages 293–339, URL: <https://doi.org/10.1093/joclec/nhu007>
- [57] Krämer J., Wohlfarth M., (2018), *Market power, regulatory convergence, and the role of data in digital markets*, Telecommunications Policy, Elsevier, vol. 42(2), pages 154-171.

- [58] OECD, (2018), *Quality considerations in digital zero-price markets*, Joint meeting of the Competition Committee and the Committee on Consumer Policy on 28 November 2018. URL: www.oecd.org/daf/competition/quality-considerations-in-the-zero-price-economy.htm
- [59] Regolamento Generale Sulla Protezione Dei Dati, GDPR, Regolamento (UE) 2016/679 del Parlamento europeo e del Consiglio del 27 aprile 2016. URL: <https://www.garanteprivacy.it/il-testo-del-regolamento>
- [60] Lucarelli S, Musolino E., Rocchi G., Ciacciarelli S., Laniado D., Espelt R., Senabre E., Morell M., Bria F, (2017), *Economic and regulatory analysis of data platforms and value creation models of the on demand economy*, DECODE, Project n. 732546
- [61] ITMedia Consulting, (2018), *L'economia Dei Dati. Tendenze di mercato e prospettive di policy*, con il contributo scientifico del Centro di Ricerca ASK Università Bocconi.
- [62] OECD, (2020), *Consumer Data Rights and Competition*, JT03461266, URL: <http://www.oecd.org/daf/competition/consumer-data-rights-and-competition.htm>.
- [63] PwC (2017), *GDPR Compliance Top Data Protection Priority for 92% of US Organizations in 2017, According to PwC Survey*, URL: <https://www.pwc.com/us/en/press-releases/2017/pwc-gdpr-compliance-press-release.html>.
- [64] Peukert C., Bechtold S.; Batikas M., Kretschmer T., (2020), *European Privacy Law and Global Markets for Data*, URL: <https://doi.org/10.3929/ethz-b-000406601>.
- [65] <https://whotracks.me/blog/gdpr-what-happened.html#fn:4>
- [66] Ezrahi A., Robertson V., (2019), *Competition, Market Power and Third-Party Tracking*, World Competition, Volume 42, Issue 1, pp. 5-19.
- [67] Bundeskartellamt e Autorité de la concurrence, (2016), *Competition Law and Data*.
- [68] Furman J., (2019), *Unlocking digital competition. Report of the Digital Competition Expert Panel*.
- [69] Bundeskartellamt, (2019), Facebook Inc. i.a. - The use of abusive business terms pursuant to Section 19 (1) GWB, B6-22/16, URL: <https://www.bundeskartellamt.de/SharedDocs/Entscheidung/EN/Entscheidungen/Missbrauchsaufsicht/2019/B6-22-16.html?nn=3591568>
- [70] Bundeskartellamt, (2019), Case Summary - Facebook, Exploitative business terms pursuant to Section 19(1) GWB for inadequate data processing, URL: https://www.bundeskartellamt.de/SharedDocs/Entscheidung/EN/Fallberichte/Missbrauchsaufsicht/2019/B6-22-16.pdf?__blob=publicationFile&v=4
- [71] Englehardt S., Narayanan A., (2016), *Online Tracking: A 1-million-site Measurement and Analysis*, Princeton Web Census.

- [72] Bundeskartellamt, (2019), *Bundeskartellamt prohibits Facebook from combining user data from different sources - Background information on the Bundeskartellamt's Facebook proceeding*, URL: https://www.bundeskartellamt.de/SharedDocs/Publikation/EN/Pressemitteilungen/2019/07_02_2019_Facebook_FAQs.pdf?__blob=publicationFile&v=6
- [73] Bundeskartellamt, (2017), *Press Release: Preliminary Assessment in Facebook Proceeding: Facebook's Collection and Use of Data from Third-Party Sources Is Abusive* URL: <https://perma.cc/HPX2-AWD9>.
- [74] European Data Protection Supervisor, (2014), *Preliminary Opinion of the European Data Protection Supervisor. Privacy and competitiveness in the age of big data: The interplay between data protection, competition law and consumer protection in the Digital Economy*.
- [75] Economides N., Lianos I., (2019), *Restrictions on Privacy and Exploitation in the Digital Economy: A Competition Law Perspective*, Centre for Law, Economics and Society (CLES) Faculty of Laws, UCL London, WC1H 0EG, CLES Research Paper Series 5/2019, URL: <https://www.ucl.ac.uk/cles/research-papers>.