UNIVERSITÀ DEGLI STUDI DI CASSINO E DEL LAZIO MERIDIONALE

CORSO DI DOTTORATO IN

ISTITUZIONI, MERCATI E COMPORTAMENTI

CURRICULUM ISTITUZIONI E AZIENDA

CICLO XXXII

LE CATEGORIE GIURIDICHE DINANZI ALLA SFIDA DEI BIG DATA

SSD: IUS/07

COORDINATORE DEL CORSO

CHIAR.MA PROF.SSA **ROSELLA TOMASSONI**

DOTTORANDO

DANIELE MARANDOLA

SUPERVISORE

CHIAR.MA PROF.SSA **IVANA MARIMPIETRI**

INDICE

CAPITOLO SECONDO

ETICA E REGOLE NELL'USO DEI DATI

CAPITOLO TERZO

DATI PERSONALI E BIG DATA: PROFILI NORMATIVI

PREMESSA

Chi, negli anni Ottanta, era un adulto o un ragazzo ricorda bene come, per prenotare un biglietto aereo o un albergo, ci si recasse presso un'agenzia di viaggi: un negozio pieno di dépliant con immagini di posti esotici, mappamondi, cartine geografiche ai muri. Allo stesso modo, chi ha svolto una tesi di laurea o di dottorato nei primi anni Novanta ricorda l'odore delle biblioteche e dei libri rilegati, le riviste da ordinare e i lunghi pomeriggi impiegati a consultare archivi cartacei, per autore e per soggetto. E poi, d'un tratto, la meraviglia dei primi computer che rendevano tutto più semplice e veloce.

Tutto quel tempo, e quelle lunghe file, per ottenere, da altri, informazioni per noi rilevanti, è oggi «risparmiato» o forse, direbbe Proust, «ritrovato». Libri e riviste accademiche sono disponibili online, facciamo ricerche in tempi rapidissimi e possiamo comprendere, grazie alle citazioni di *Google Scholar*, quali filoni di ricerca si siano attivati a partire da altre ricerche di base.

Ma la riduzione del nostro tempo di «ricerca» non è l'unico vantaggio che ci ha regalato l'era digitale. Abbiamo eliminato molte di quelle intermediazioni che hanno impegnato il tempo di attenzione di intere generazioni. In un certo senso, abbiamo restituito al mercato, e all'autoorganizzazione, mansioni e professioni che una volta rendevano centralizzata, in vario grado, l'informazione rilevante. Interroghiamo quotidianamente, per diverse ore al giorno, il *Web* e ne otteniamo quasi sempre le informazioni che cerchiamo.

Piattaforme come *Amazon Prime* ci permettono di avere, in uno o due giorni, prodotti acquistati online, magari da un paese lontano, interrogando siti specializzati che ci

offrono le migliori occasioni e accorciano le distanze.

Ma l'efficienza della rivoluzione digitale si manifesta anche nelle «raccomandazioni», nei «suggerimenti» algoritmici che riceviamo dalle pubblicità dei motori di ricerca o evidenziate nei siti di *e-commerce*: occasioni che potrebbero interessarci, come hanno interessato altri, simili a noi, prima di noi. Il tempo che risparmiamo non è solo quello che spenderemmo a cercare informazioni, ma anche quello che impiegheremmo a capire quali informazioni cercare e perché. Questa efficienza algoritmica è il risultato di un patto implicito: per avere le migliori informazioni dal Web, anche noi dobbiamo rivelare le informazioni che ci riguardano.

I *big data* e gli algoritmi che li «macinano» aprono la frontiera dell'intelligenza artificiale, definita dall'Unione europea come «l'insieme di quei sistemi che mostrano un comportamento intelligente, per cui, analizzando l'ambiente, possono svolgere vari compiti con un certo grado di autonomia per conseguire specifici obiettivi». L'organizzazione produttiva dell'industria manifatturiera sta cambiando forma, ricorrendo, in modo sempre più accentuato, alla decentralizzazione produttiva da un lato e all'automazione dei processi e alla robotica dall'altro, funzioni rese possibili dalla fruizione di moltissimi dati in tempo reale. Tutto questo cambia la struttura organizzativa sia delle imprese sia dei mercati. La riduzione dei costi e l'aumento della scala produttiva si traducono in prezzi minori per i consumatori. L'imprenditore oggi non è soltanto l'organizzatore di una struttura verticale di produzione e di controllo, ma può diventare un collettore «intelligente» di input e di attività decentrate e parcellizzate, svolte in luoghi diversi a minor costo, per realizzare prodotti e servizi destinati al mercato globale.

Come dimostrano molti studi dell'Ocse, anche il rapporto di lavoro è destinato a

cambiare. Da diversi anni abbiamo inaugurato un lungo periodo di transizione nel quale vecchi mestieri non sono più richiesti, mentre nuove professionalità non si sono ancora sviluppate. Il che genera una disoccupazione crescente in assenza di misure adeguate di conversione e formazione.

Siamo circondati dai dati, vi siamo immersi. Ogni minuto vengono scambiate, nel mondo, milioni di informazioni. Solo per fare qualche esempio in continuo aggiornamento, ogni 60 secondi, su *Facebook*, vengono creati 3,3 milioni di post, pubblicati 510.000 commenti e aggiornati 293.000 stati; su *Twitter* vengono inviati 470.000 *tweet*; su *WhatsApp* vengono scambiati 38 milioni di messaggi; su Google vengono effettuate 3,8 milioni di ricerche. In altre parole, ci rapportiamo con un flusso continuo e ininterrotto di informazioni, notizie reali e false che coesistono e possono confondersi.

Accanto a queste statistiche ci sono episodi che segnano, nella nostra memoria, una rivoluzione dei dati. Nell'aprile del 2017, a circa sei mesi dall'elezione di Donald Trump, Chuck Todd su Nbs News affermava: «i *big data* hanno ormai distrutto la politica americana». Appena un mese dopo, una celebre copertina dell'«Economist» sentenziava: «La risorsa più preziosa al mondo non è più il petrolio, sono i dati». Un anno dopo, il 10 aprile 2018, il fondatore e Ceo di Facebook, Mark Zuckerberg veniva chiamato a testimoniare sulla tutela della privacy digitale nel social media e sull'uso dei dati raccolti e profilati, dopo il clamore del caso *Cambridge Analytica*, davanti alle commissioni congiunte per il commercio e la giustizia del Senato americano. Dal 25 maggio 2018 è direttamente applicabile, in tutti gli Stati membri dell'Unione Europea, la *General Data Protection Regulation* (Gdpr) sulla protezione delle persone fisiche rispetto al trattamento e alla libera circolazione dei dati.

Politica, economia, società, privacy digitale: i *big data* entrano dappertutto, in ogni sfera della nostra vita, pubblica e privata. E il dibattito circa la loro estrazione e il loro impiego ha assunto ormai le caratteristiche di una vera e propria emergenza. Discorsi in cui si alternano i pericoli e le opportunità nell'uso dei dati che, consapevoli o meno, immettiamo in rete o rilasciamo nella nostra vita quotidiana.

Scandali riguardanti la violazione dei dati personali in rete - da quello, nel 2006, relativo ai dati di ricerca di 650.000 utenti di Aol, a quello, del 2013, legato al programma Prism dell'agenzia statunitense Nsa, fino al più recente caso, come visto, Cambridge Analytica - hanno corroborato i motivi di quest'apprensione e hanno focalizzato il dibattito mediatico e istituzionale sulla «questione dei dati». In origine, il dibattito ha riguardato l'aspetto di riservatezza dei dati personali.

Oggi alla tutela della privacy si affiancano due nuove grandi tematiche: da un lato, l'analisi del vantaggio competitivo dell'uso esclusivo dei dati a fini di profilazione commerciale da parte delle grandi piattaforme globali, a partire dai *big five* Google (anche con YouTube), Amazon, Facebook (anche con WhatsApp, Instagram e Messenger), Apple, Microsoft (anche con Skype e Linkedin); dall'altro, la crescente preoccupazione circa l'impatto del rilascio e dell'uso di dati a fini di marketing politico indiretto, anche per il tramite di sofisticate strategie di disinformazione che sembrano aver già interessato diversi importanti appuntamenti elettorali, in varie parti del mondo, seppure con esiti tutti ancora da determinare.

In Italia, secondo il rapporto *Digital in 2018 Global Overview* redatto da *We Are Social*, spendiamo mediamente sei ore al giorno in rete davanti al desktop del nostro computer e quasi due utilizzando una piattaforma social media. A livello mondiale, l'utilizzo di Internet da postazione mobile (telefoni cellulari, tablet, smart Tv) ha già

superato l'accesso da postazione fissa, mentre in Italia si registra ancora un'equa ripartizione tra le due modalità. La disponibilità di servizi di connettività mobile ad alta capacità (le generazioni 4G) ha fatto esplodere la domanda di contenuti e aumentato il tempo dedicato alla connessione. Significa più tempo di attenzione, maggiori informazioni fruite e maggiori dati rilasciati nell'ecosistema digitale. L'Osservatorio sulle comunicazioni dell'Autorità per le garanzie nelle comunicazioni (Agcom), mostra che, con la definitiva affermazione delle connessioni 4G, il traffico dati da telefonia mobile è aumentato del 56% e i consumi unitari sono passati da 1,84 a 2,76 Giga/mese ad utenza, con una crescita del 49,5% rispetto all'anno precedente. Il numero di carte Sim con accesso ad Internet è di 52,2 milioni di unità, pari al 63,9% dell'intera base clienti, quasi il doppio rispetto al 2012, quando questa tipologia di Sim rappresentava il 27,8% del totale.

Nei prossimi sei anni, il traffico dati da mobile aumenterà fino a cinque volte e registrerà un forte incremento della domanda di servizi che supportano *streaming video* di grandissima qualità, applicazioni e giochi di realtà aumentata, automazione industriale, robotica avanzata, operazioni da remoto, guida autonoma e così via.

Ciò sarà anche il risultato delle connessioni mobili di «quinta generazione» (5G) rese possibili da nuove risorse frequenziali, nuovi terminali, nuovi apparati e nuove antenne capaci di trasmettere una grande mole di dati, anche in mobilità, in «tempo reale».

I social network sono definitivamente divenuti parte integrante della dieta informativa quotidiana dei cittadini in Italia e nel mondo. Le piattaforme online, in generale, basano il proprio business sull'estrazione, il trattamento e l'elaborazione di informazioni e dati da profili personalizzati, la cui disponibilità aumenta in relazione alla crescente intensità d'uso della rete da parte di cittadini, consumatori, imprese e

istituzioni.

Le piattaforme online sono diventate così i nuovi leader mondiali nel settore della pubblicità, sottraendo risorse pubblicitarie crescenti ai media tradizionali e rappresentando ormai il veicolo distributivo principale per l'accesso e la diffusione dell'informazione in rete. Ma cosa sono i *big data*? E perché è così rilevante comprenderne il ruolo nella società digitale?

I *big data* stanno cambiando il nostro mondo. In molti modi. C'è chi parla di *singolarità,* una parola che rappresenta un drastico big bang, un punto di non ritorno per l'umanità, con la perdita del controllo e della capacità di apprendere o di conoscere e gestire la tecnologia, come l'enorme mole di dati che ci circondano, consegnandosi al dominio dei robot. E c'è chi, come il filosofo Luciano Floridi, invita, invece, a superare queste visioni distopiche e a disegnare una nuova *governance* e una nuova ecologia che rimettano al centro l'uomo, ma con un nuovo sé *onlife*, per governare e indirizzare la *quarta rivoluzione*.

In questo lavoro affronteremo soprattutto i cambiamenti in corso nei rapporti di *potere economico* sui mercati intermediati dalle piattaforme digitali globali, nella relazione tra concorrenza e innovazione, nonché nei possibili rapporti di *potere politico*, in particolare con la capacità di influenzare la formazione dell'opinione pubblica, specie alla vigilia di importanti appuntamenti elettorali, a seguito di strategie di *micro-targeting* politico e di disinformazione, spesso alimentate da temi divisivi e da campagne di discriminazione o da espressioni d'odio (*hatespeech*) nei confronti di gruppi o categorie di persone. Non tratteremo, se non marginalmente, il tema della privacy, in particolare di quella detta «digitale», affrontato in molti altri volumi.

Questo elaborato non si occupa della protezione del dato personale, ma, cambiando la

prospettiva, segue un approccio economico che mette al centro la transazione digitale dei dati e le sue regole, dallo «scambio implicito» del dato alla costruzione del «mercato dei dati», temi sui quali sappiamo ancora troppo poco.

La profilazione e lo scambio dei dati offrono nuove opportunità, ma pongono anche alcune rilevanti domande. La cessione del dato è una transazione economica che certifica la natura «proprietaria» del dato o è solo una manifestazione del consenso che ne «delega» il trattamento, pur entro certi limiti? Siamo consapevoli di partecipare ad una transazione economica nella quale stiamo alienando un bene di cui altri, a vario titolo, estrarranno il valore? Ha ancora senso, sotto il profilo dell'estrazione e dell'utilizzazione economica, la distinzione tra dati personali e non personali? Com'è accaduto che siamo stati (auto) esclusi da quel «mercato dei dati» che pure la nostra attenzione ha generato, nello scambio implicito tra attenzione e servizi in gran parte «gratuiti»? Dove si crea e verso dove si indirizza il valore del dato? Come cambiano i mercati e i diritti di proprietà in transazioni basate sul valore del dato? C'è bisogno di nuova regolazione, accanto alla privacy e all'antitrust, o basta la concorrenza a disciplinare innovazione e potere di mercato?

Allo stesso modo, i nostri dati e la nostra profilazione generano un «valore politico» nel mercato del consenso elettorale. Riceviamo stimoli, propaganda e messaggi profilati volti a influenzare la formazione delle nostre opinioni o a rafforzare e polarizzare la visione del mondo che la nostra profilazione rivela. Anche in questo caso, senza esserne pienamente consapevoli. Che ne è allora del pluralismo nella intermediazione informativa basata sulla selezione degli algoritmi? Come possiamo conciliare, in un percorso equilibrato, il valore commerciale dell'informazione e il rispetto di diritti individuali e collettivi fondamentali quali la privacy digitale, la tutela

della *cybersecurity*, la tutela della concorrenza e le garanzie del pluralismo informativo?

Si cercherà, nel corso della presente trattazione, di offrire una mappa per orientarsi e per comprendere alcune sfide profonde che abbiamo di fronte.

CAPITOLO PRIMO

I DATI: UN PATRIMONIO PER LA CREAZIONE DI VALORE

1. Definizione

Ne parlano tutti, è vero, eppure non vi è una definizione univoca di *big data*. Secondo l'Unione europea i *megadati* sono «grandi quantità di tipi diversi di dati prodotti da varie fonti, fra cui persone, macchine e sensori»[1].

Il «big» dei dati fa riferimento ad alcune caratteristiche fondamentali: la *velocità*, la *varietà* e il *volume* dei dati raccolti e processati. È da queste tre caratteristiche che si genera - come vedremo - la quarta *v*, il *valore* dei dati: rilevazioni su meteo, clima e ambiente, immagini satellitari, immagini e video digitali, registrazioni di operazioni, segnali geo-localizzati (Gps), protocollo Internet (Ip), ricerche online, messaggi su social network, acquisti online, informazioni sanitarie, occupazionali, dati personali cosiddetti «strutturati» quali le informazioni riguardanti una persona (nome, foto, indirizzo email, estremi bancari) e così via. La gran parte di questi dati è di solito «non strutturata», ossia viene acquisita e immagazzinata secondo criteri che differiscono da quelli dei tradizionali database ben organizzati (quali quelli «relazionali»)[2].

Ciò che è rilevante è il processo di lavorazione e aggregazione dei dati e di questi ultimi con gli algoritmi. I *big data* servono a migliorare l'algoritmo e, a sua volta, l'uso dell'algoritmo da parte di ciascuno di noi genera nuovi dati, e così via, insegnando all'algoritmo come migliorare e, persino, come «imparare ad imparare meglio». Si pensi all'insieme di informazioni che ciascun utente genera navigando in rete (ad esempio, il motore di ricerca che interroghiamo o i suggerimenti che ci dà Amazon quando acquistiamo un prodotto), lasciando una vera e propria impronta individuale (*digital footprint*). Questi dati grezzi prendono il nome di *data exhaust*, si tratta di numerosissime informazioni (*cookies*, file temporanei, *logfiles*, parole digitate, ecc.) acquisite a grandi velocità e composte dai formati più vari.

---

[1] ANDERSON C., *The End of Theory: the Data Deluge Makes the Scientific Method Obsolete*, «Wired», 2008, pp. 34 ss.

[2] *Ibidem*, p. 37.

L'avvento dei *big data* determina un nuovo approccio al trattamento dei dati, capovolgendo la relazione tra domanda di ricerca e risultati, al punto che alcuni studiosi, come Chris Anderson, si sono chiesti se non siamo giunti «alla fine della teoria»[3]: a che servono, infatti, le teorie se sono ormai i dati a rivelarci correlazioni e causazioni? Il tema è in realtà complesso perché le teorie poste a verifica empirica hanno l'ambizione della generalizzazione del risultato e di indagare regolarità non contingenti[4].

Non c'è dubbio tuttavia che oggi gli algoritmi, grazie al trattamento di grandi masse di dati, possono rivelare relazioni (*correlation insights*) tra scelte, comportamenti, azioni, gusti, e aiutare così a costruire «modelli», altamente predittivi, di domanda e offerta di prodotti, servizi, contenuti di vario genere (inclusi i contenuti culturali e politici). Ecco perché i dati sono, al contempo, un «input», cioè uno strumento d'indagine utile per soddisfare preferenze e domanda di prodotti e servizi, ma anche un «prodotto» in sé, capace di generare valore autonomo, per esempio con la pubblicità personalizzata e la commercializzazione di prodotti e servizi.

Dal momento che l'efficienza degli algoritmi dipende dal volume dei dati, le piattaforme digitali globali, che accedono a centinaia di milioni di profili di utenti, rivestono un ruolo fondamentale in questo processo, al punto, come ha rilevato ironicamente Scott Galloway[5], da poter sapere, potenzialmente, di ciascuno di noi, più di quanto sappiamo di noi stessi. L'utilizzo dei *data exhaust* ha consentito a Google di perfezionare sempre più il proprio motore di ricerca, vincendo la concorrenza su Yahoo!, Bing e altri. Al tempo stesso, la nostra impronta digitale ci insegue: improvvisamente, dopo aver visitato un sito Web di un negozio di mobili, ci ritroviamo la pubblicità di quel negozio - solo per noi - tra le inserzioni di un quotidiano d'informazione online, tra i post di Facebook, nelle pubblicità di Google. Gli algoritmi lavorano sempre e non si distraggono mai[6].

Un grande volume di dati e dunque utile per l'*analitica previsionale* e per migliorare la nostra vita, coordinando al meglio le azioni collettive. Si pensi all'organizzazione

---

[3] Cfr. ***ANDERSON C.,*** *The End of Theory: the Data Deluge Makes the Scientific Method Obsolete*, op. cit., p. 41.
[4] *Ibidem*, p. 36.
[5] SCOTT G., *The Four. I padroni: Il DNA segreto di Amazon, Apple, Facebook e Google*, Hoepli, Milano, 2018, pp. 4 ss.
[6] *Ibidem*, p. 8.

dell'assistenza sanitaria, al controllo del traffico stradale, alla raccolta dei rifiuti, alla programmazione di energia elettrica e così via. Talvolta ciò genera preoccupazioni per le nostre libertà, per la nostra privacy, persino per la nostra sicurezza.

Per questa ragione, ogni analisi sui dati deve tener conto del processo aggregativo dell'algoritmo e, allo stesso tempo, della natura e degli obiettivi, economici e sociali[7].

Il fenomeno "Big Data", nato attorno al 2011, ha ormai superato la fase in cui si trattava di una *buzzword*, cioè una parola in voga, alla moda, che è stata utilizzata più per promuovere prodotti o servizi informatici che per risolvere problematiche legate all'estrazione di valore da una nuova categoria di dati[8].

Oggi, infatti, la disponibilità di tecnologie sia *open source* sia proprietarie e la disponibilità di piattaforme *cloud* sono fattori che concorrono a rendere più semplice e meno costosa l'attivazione di processi volti a sfruttare le opportunità che i big data offrono[9].

Di fatto, anche in Italia sono molte le aziende, a partire dai grandi gruppi bancari, che si sono dotate o si stanno dotando di sistemi quali Hadoop e Spark per la raccolta sistematica e l'elaborazione di grandi volumi di dati strutturati e non. Gli obiettivi di tale sforzo sono molteplici: da un lato vi è la riduzione dei costi che tecnologie di questo tipo possono apportare in termini di *storage* e di licenze software; dall'altro vi è la possibilità di analizzare dati con una granularità e una profondità storica che con altri strumenti (gli RDBMS, i *database* relazionali) non era pensabile, se non a costi elevatissimi e sproporzionati rispetto ai benefici che ne derivavano[10].

Le analisi sono condotte in primo luogo con strumenti di analisi descrittiva, già presenti in azienda, ma che oggi si sono dotati di connettori per interfacciarsi con le nuove basi di dati. Tuttavia, il valore aggiunto proviene dall'impiego di tecniche di analisi avanzata, in particolare dalle analisi predittive e prescrittive, in grado di creare vantaggi competitivi. Le analisi predittiva e prescrittiva si differenziano dall'analisi di tipo descrittivo, poiché quest'ultima è rivolta al passato e osserva ciò che è accaduto,

---

[7] DE MAURO A., *Big Data Analytics. Analizzare e interpretare dati con il machine learning*, Apogeo, Milano, 2019, pp. 86 ss.
[8] *Ibidem*, p. 88.
[9] MAYER-SCHONBERGER V. - CUKIER K., *Big data. Una rivoluzione che trasformerà il nostro modo di vivere e già minaccia la nostra libertà*, Garzanti, Milano, 2013, pp. 23-26.
[10] *Ibidem*, p. 29.

misurandone gli effetti, mentre le altre due tipologie sono proiettate nel futuro[11].
Esse infatti cercano di anticipare gli eventi, consentendo al management di prendere decisioni tempestive o, meglio, anticipate, rispetto a un dato accadimento.

2. Caratteristiche dei Big Data: le "5 V"

I big data sono generalmente definiti come quei dati che presentano una o più delle seguenti cinque caratteristiche, chiamate le "5 V"[12].
Nel 2001, Doug Laney, allora vice presidente e Service Director dell'azienda Meta Group, descrisse in un report il *Modello delle 3V* relativo alle 3V dei Big Data: *Volume*, *Velocità* e *Varietà*[13]. Un modello semplice e sintetico per definire dei nuovi dati, generati dell'aumento delle fonti informative e più in generale dall'evoluzione delle tecnologie.
Oggi il paradigma di Laney è stato arricchito dalle variabili di *Veridicità* e *Variabilità* e per questo si parla di "5V" dei Big Data.
Le "3 V" dei Big Data: quali sono?
Si definiscono Big Data quei dati che abbiano almeno una delle seguenti caratteristiche[14]:

- Volume: cioè le quantità elevate di dati (a partire da decine di terabyte in su). I dati generati automaticamente da macchine (sensori, DCS - *Distributed Control System*, strumenti scientifici) e quelli relativi a transazioni bancarie e movimenti sui mercati finanziari possono assumere volumi imponenti, soprattutto se considerati al loro massimo livello di granularità. Ogni giorno, in moltissime attività della nostra vita quotidiana, generiamo dati. Con volume si fa riferimento quindi a quest'ingente massa di informazioni, che non è possibile raccogliere con tecnologie tradizionali. Questo volume di dati è in continua crescita, gli analisti internazionali stimano che la produzione di dati nel 2020 sarà 44 volte maggiore di quella del 2009. Proprio per questo è difficile identificare un valore limite al di sopra del quale si può parlare di Big Data.

[11] *Ibidem*, p. 34.
[12] REZZANI A., *Big data. Architettura, tecnologie e metodi per l'utilizzo di grandi basi di dati*, Apogeo, Milano, 2013, p. 41.
[13] LANEY D., *3D Data Management*: *Controlling Data Volume, Velocity, and Variety*, Meta Group, 2001, pp. 11 ss.
[14] *Ibidem*, p. 16.

Per adesso, consideriamo la soglia di più di 50 Terabyte o volumi di dati che crescono più del 50% annuo.

- Velocità: indica la rapidità con cui i dati sono prodotti. Pensiamo al mondo dell'*Internet Of Things* e dei sensori, che sono in grado di generare dati con una velocità elevatissima. I dati nascono e vengono acquisiti sempre più rapidamente. Basta pensare alla proliferazione di dispositivi dotati di sensoristica capace di raccogliere dati in tempo reale. La sfida, con cui le aziende sono chiamate a confrontarsi, è la necessità non solo di raccogliere questi dati ma anche analizzarli in tempo reale, per poter prendere decisioni di business con la maggiore tempestività possibile.
- Varietà: riguarda la diversità dei formati, delle fonti e delle strutture. Inoltre, alcuni dati possono anche non avere una struttura. "*More isn't just more. More is different.*" – così scriveva Chris Anderson sul magazine Wired, era il 2008. Con varietà si fa riferimento proprio alle differenti tipologie di dati oggi disponibili, provenienti da un numero crescente di fonti eterogenee. Non solo sistemi transazionali e gestionali aziendali, ma anche sensori, social network, open data. Dati sia strutturati che non, sempre di più non solo dati interni all'organizzazione ma anche acquisiti esternamente.

Perché oggi parliamo di 5V?

Nei primi anni 2000 si definivano i Big Data con tre parole: volume, velocità e varietà. Con il passare degli anni, mentre il termine perdeva la sua aurea fantascientifica per diventare sempre più concreto e applicabile nelle aziende, ci si è chiesto se non vi fossero altre caratteristiche da mettere in risalto.

Hanno arricchito il modello due nuove V, volte a definire come questi nuovi dati debbano essere utilizzati[15]:

- Veridicità: tra gli addetti al settore, alcuni usano dire *"Bad data is worse than no data"*. I dati devono essere affidabili, raccontare il vero. Con i Big Data questa sfida è ancora più difficile da affrontare: cambiano le tecnologie di gestione dei dati, cambia la velocità con la quale i dati vengono raccolti e

[15] MAYER SCHÖNBERGER V. - CUCKIER K., *Big Data: una rivoluzione che trasformerà il nostro modo di vivere e già minaccia la nostra libertà*, Garzanti, Milano, 2013, pp. 9 ss.

aumentano le fonti. La qualità e l'integrità delle informazioni rimane però un pilastro imprescindibile per dar vita ad analisi che siano utili e affidabili.

- Variabilità: molti più dati, in diversi formati e provenienti da diversi contesti. La mutevolezza del loro significato è un aspetto da tenere in considerazione nel momento in cui i dati vengono interpretati. Ancor di più, se a farlo è un utente che lavora in una linea di business e non solo il *data scientist*.

I Big Data sono stati definiti negli ultimi anni come il nuovo petrolio o il nuovo oro, ovvero una fonte inestimabile di valore. È esattamente così. Ma limitarsi a raccogliere i dati, pur sfruttando le migliori tecnologie disponibili sul mercato, non garantisce di avere informazioni e soprattutto di estrarre conoscenza[16].

Parlare di dati, informazioni e conoscenza vuol dire parlare di aspetti in relazione tra loro ma differenti. Per definizione un dato è una rappresentazione codificata di un'entità, di un fenomeno, di una transazione, di un avvenimento. L'informazione è il risultato di un processo di analisi del dato, spesso ha significato soltanto per colui che opera nel dominio di generazione del dato. La conoscenza si ottiene quando una persona utilizza le informazioni per prendere decisioni e realizzare azioni, quando le informazioni vengono utilizzate per esser messe "in pratica"[17].

Per attuare questo processo, e far sì che i Big Data possano essere trasformati in informazioni da utilizzare nei processi aziendali costruendo conoscenza per migliorare le performance, sono necessari strumenti di Analytics. Oltre il modello delle 5V, quindi, è doveroso considerare una ulteriore V, ossia la sesta: il *valore*. Qui sono le metodologie di *Big Data Analytics* ad essere fondamentali, attraverso il loro utilizzo un'azienda può estrarre valore, ovvero prendere decisioni più informate, tempestive e consapevoli, dal vasto mondo dei Big Data[18].

Quando parliamo di tecnologie tradizionali intendiamo essenzialmente i database relazionali (RDBMS, *Relational Database Management Systems*), per quanto riguarda le basi dati e gli strumenti di analisi descrittiva che li utilizzano come sorgenti, ma anche quei *tool* di analisi predittiva e *data mining* che mostrano i loro limiti al crescere dei volumi di dati o al venir meno di strutture tabellari. La definizione basata non solo

[16] *Ibidem*, p. 14.
[17] FLEISH E., *What is the Internet of things? An economic perspective*, Economics, management and financial markets, 2010, pp. 54 ss.
[18] *Ibidem*, p. 57.

sui limiti tecnologici, ma anche sulla convenienza economica dovrebbe far immediatamente comprendere al lettore che non è strettamente necessario possedere moli di dati dell'ordine dei *petabyte* per pensare all'adozione di sistemi big data (Hadoop e Spark, per esempio)[19].

È ovvio che la componente del *volume* gioca un ruolo importante sia nella diminuzione della convenienza economica delle tecnologie tradizionali, sia nel raggiungimento dei loro limiti tecnici. Anche la diversità (o l'assenza) di formati può rendere poco adatto un database relazionale all'analisi di certi dati. Quanto alla *velocità*, essa produce criticità soprattutto nelle fasi di acquisizione (*data ingestion*) e di salvataggio nel database[20].

Il termine *big data* identifica dunque sia i dati con le caratteristiche sopra descritte, sia le tecnologie con cui si possono risolvere i problemi (tecnici o economici) di *data ingestion*, di conservazione e di analisi dei dati.

3. Il valore economico dei dati

Un famoso *spot*, utilizzato per oltre quarant'anni da L'Oréal, recitava *Because you're worth it* («Perché tu vali»)[21]. Bellissime modelle pubblicizzavano prodotti cosmetici, ma il messaggio suggeriva che il valore non risiedeva tanto nei prodotti, quanto nella persona che li usava e che li meritava: una donna, ancor prima che una modella. Erano gli anni Settanta e quello spot fu un cambio di prospettiva di enorme successo.

Oggi si potrebbe dire: «noi valiamo» (anche) perché i nostri dati valgono. Per due ragioni su tutte: *a*) perché i dati permettono di profilare la nostra domanda *individuale* di consumo di servizi e prodotti, rendendo assai efficaci forme di pubblicità e di commercializzazione personalizzata, aumentando la probabilità di vendita; b) perché i dati consentono agli algoritmi di migliorare sé stessi, man mano che nuovi dati sono analizzati, e di stimare, così, la domanda *aggregata* o media di consumo di servizi e prodotti, indicando in tempi assai rapidi le evoluzioni delle preferenze, i bisogni del mercato, le opportunità di investimento e di innovazione e così via. Il che permette di sviluppare al massimo il rendimento degli investimenti pubblicitari attratti dalle grandi

[19] WHITE T., *Hadoop: The Definitive Guide*, 4th Edition, O'Reilly Media, 2015, pp. 48 ss.
[20] *Ibidem*, p. 55.
[21] VALENTE P. - IANNI G. - ROCCATAGLIATA F., *Economia digitale e commercio elettronico*, Ipsoa, Milano, 2015, pp. 32 ss.

piattaforme online. Il valore dei dati aumenta con il loro volume e la loro varietà[22]. E per aumentare il valore dei dati il mercato e l'industria si sono evoluti con modelli di business nuovi, in modo tale da stimolare in ciascuno di noi la massima intensità di rivelazione di dati. Come? Con l'avvento del paradigma del *free,* che nella lingua inglese significa «libero», ma anche «gratuito». Una straordinaria combinazione semantica che tuttavia alimenta, e non poco, la confusione sul concetto di «libertà di scelta». Nell'ecosistema digitale possiamo accedere a moltissimi servizi *gratuitamente* e dedicare molto del nostro tempo di attenzione ad essi. Tutto ciò che dobbiamo fare è semplicemente dare nome, cognome, indirizzo mail, numero di telefono, una password con la massima garanzia - rafforzata dalla normativa a tutela della privacy - circa l'utilizzo di quei dati che rilasciamo. E il gioco è fatto. Ma quello che ci sembra un banale strumento per ottenere un *libero* accesso è in realtà il vero bene, il cui scambio regge la transazione commerciale sottostante. Lo *scambio implicito*, per tutta questa gratuità di servizi, è con la nostra attenzione, con il rilascio di dati che permetteranno poi promozioni e pubblicità personalizzate per i nostri bisogni. A questo scambio implicito corrisponde un *mercato implicito*, quello dei dati, del quale sappiamo ancora troppo poco[23]. Come spesso si ripete in questi casi, il *prodotto siamo noi:* l'informazione *rivelata* sulle scelte che abbiamo compiuto, sulla nostra disponibilità a pagare, sulla frequenza ad acquistare, sul tempo dedicato alla fruizione del prodotto o alla sua ricerca e così via. La nostra attenzione da un lato produce dati che servono anche a profilarci, dall'altro è la destinataria di informazioni (pubblicitarie e non solo) profilate per noi.

In una recente indagine, Agcom ha analizzato un *dataset* su oltre un milione di applicazioni presenti su Google Store, mostrando, con un'analisi econometrica, come le *app* gratuite richiedano la cessione di un numero significativamente maggiore di dati individuali rispetto a quelle a pagamento[24]. Nel caso delle *app* gratuite, vengono cioè richiesti permessi sul rilascio di dati ultronei che non incidono direttamente sul funzionamento del servizio, come quando una «torcia» ci chiede accesso all'agenda, alla posizione, alla telecamera presenti nel nostro *smartphone*. Dunque il *free* di molti

[22] *Ibidem*, p. 40.
[23] REZZANI A., *Big data. Architettura, tecnologie e metodi per l'utilizzo di grandi basi di dati*, op. cit., p. 49.
[24] GARANTE PER LA PROTEZIONE DEI DATI PERSONALI, *Big Data: Agcom, Antitrust e Garante privacy avviano indagine conoscitiva*, 2017.

servizi online non è «libero», perché condizionato ad uno scambio, e non è «gratuito» perché è pagato con i nostri dati[25].

Ma come viene creato il *valore dal dato?* Come si struttura il *mercato dei dati?* Come si distribuisce il *valore del dato?*

Descrivere i dati come un bene economico può far sgranare gli occhi a quanti vedono, giustamente, nel dato personale una caratteristica intimamente connessa con i diritti della persona. L'analisi economica del dato, tuttavia, è di ordine pragmatico e prescinde, ovviamente, da ogni considerazione di ordine etico e giuridico circa lo sfruttamento economico di informazioni di natura personale e il significato della loro tutela nella sfera dei diritti fondamentali dell'individuo[26].

Se (indipendentemente dal grado di apertura di questo mercato «implicito») seguiamo un approccio pragmatico e cerchiamo di capire dove stanno gli interessi e gli incentivi privati (*follow the money*), ci rendiamo subito conto che non basta affermare il principio di un dato personale sacro e inviolabile, per essere sicuri che qualcuno non ci abbia già costruito sopra un business o un mercato.

Secondo la riflessione di Alessandro Acquisti della *Carnegie Mellon University*, tra i primi economisti ad occuparsi di dati personali da un diverso punto di vista, il dato (personale e non) è di fatto un bene economico perché ha una domanda e un'offerta e perché produce valore[27].

Il problema è che i dati sono un bene economico, ma sono caratterizzati da una pervasiva ambiguità, o incompletezza, circa la loro natura di diritti proprietari. Il diritto di proprietà attribuisce al proprietario la titolarità di un insieme di usi (*entitlements*), i quali possono essere oggetto di specifica transazione economica[28]. Viene dunque da chiedersi se, quando rilasciamo un «consenso» all'uso del dato, stiamo assistendo o meno a un passaggio di proprietà su quello specifico uso.

Per molti studiosi non è affatto così: i dati personali ceduti sarebbero una mera «delega», funzionale unicamente a identificare la persona per permettere, grazie ad un certo uso «limitato» del dato, l'erogazione del servizio. Il servizio, poi, non sarebbe

---

[25] *Ibidem*.

[26] ARICKER M. - MCGUIRE T. - PERRY J., *Harvard Business Review: Five Roles You Need on Your Big Data Team,* 2013, pp. 76 ss.

[27] *Ibidem*, p. 78.

[28] HARTMANN P. - ZAKI M. - FIELDMANN N. - NEELY A., *University of Cambridge: Big Business? A taxonomy of Data-driven Business Models used by Start-up firms*, marzo 2014, pp. 43 ss.

ceduto «in cambio del dato», ma offerto gratuitamente al fine di attrarre l'attenzione del maggior numero di soggetti cui veicolare la pubblicità pagata da terzi su un altro versante del mercato. Il dato personale, in questa lettura, altro non è che un bene inalienabile, privo di natura «proprietaria», la cui circolazione è vietata, così che coloro cui viene delegato l'uso *devono* detenerlo in via esclusiva[29].

Accanto a questo tradizionale approccio, si va affermando una ricostruzione opposta, secondo la quale, invece, proprio la previsione della condizione di un consenso per l'accesso al dato ne rivelerebbe la natura proprietaria *de facto* e, entro determinati limiti, la sua «alienabilità». Ad esempio, come ci ha insegnato Guido Calabresi, professore emerito a Yale e giudice dell'*US Court of Appeals for the Second Circuit*, l'obbligo di ottenere il consenso per l'accesso a determinati beni caratterizza la tutela inibitoria (*property rule*)[30], cioè la regola di protezione forte del diritto di proprietà rispetto alla più debole tutela risarcitoria (*liability rule*) che assicura al proprietario solo un compenso, nel caso di accesso di terzi senza consenso[31].

Riconoscere, a chi generi un dato, la proprietà dello stesso significherebbe rendere esplicita, a questo punto su un vero e proprio mercato dei dati, la transazione tra chi genera il dato e chi lo acquisisce pagando un prezzo (oppure offrendo un servizio o corrispondendo un'utilità) per determinati usi. Ma oggi questo tipo di transazioni *esplicite*, almeno tra l'originario produttore del dato e chi lo acquisisce per fini industriali, non avviene sul mercato.

Diventa allora importante comprendere che la questione della cessione del dato, e della sua valorizzazione economica, non rileva solo per la tutela della *privacy*, ma anche al fine della costruzione giuridica, oltre che economica, di un vero e proprio mercato trasparente dei dati[32].

L'ambiguità, tra «delega» e «proprietà», circa la natura del consenso al rilascio dei dati si comprende forse meglio ricorrendo a questo gioco di parole: il proprio dato è un dato proprio? In altri termini: le informazioni che ci riguardano, per il fatto stesso che riguardano noi (e che potremmo aver generato con il nostro comportamento), sono di

---

[29] *Ibidem*, p. 45.

[30] CALABRESI G., *Property Rules, Liability Rules, and Inalienability: One View of the Cathedral*, Harvard Law Review, Harvard, 2010, pp. 143 ss.

[31] *Ibidem*, p. 147.

[32] MAYER-SCHONBERGER V. - CUKIER K., *Big data. Una rivoluzione che trasformerà il nostro modo di vivere e già minaccia la nostra libertà*, op. cit., p. 55.

nostra proprietà? E se non sono di nostra proprietà, sono di proprietà di qualcun altro o sono un bene pubblico? E se da quella informazione un terzo genera un valore privato, coloro che quella informazione hanno prodotto, se non altro per esserne l'oggetto, hanno diritti sulla (re)distribuzione, almeno in parte, di quel valore?[33]

Si tratta di questioni dirimenti perché su questa ambiguità si è costruito un paradosso: le informazioni che ci riguardano e che rilasciamo nella nostra impronta digitale hanno le caratteristiche di un bene pubblico ma, a differenza di ogni bene pubblico, vengono poi fatte proprie in via esclusiva da soggetti terzi che ne estraggono un valore economico privato. È un modello di business che produce anche significativi benefici sociali, ma che è costruito su un'inedita e ambigua caratterizzazione e definizione dei diritti di proprietà[34].

L'informazione è un bene pubblico in quanto: *a*) il suo consumo (o accesso) da parte di un soggetto non implica l'impossibilità per un altro soggetto di consumarlo nello stesso momento (c.d. «assenza di rivalità nel consumo»); *b*) una volta prodotta, è difficile impedirne la fruizione ai soggetti che non hanno pagato per averla («non escludibilità nel consumo»). Le informazioni possono essere, in teoria, «consumate» all'infinito e da una molteplicità di soggetti e la capacità di escluderne l'accesso a terzi dipende da tanti fattori. Per esempio, l'invenzione del diritto d'autore attribuisce una titolarità, temporanea, nella forma di un pieno diritto di proprietà (*property rule*), all'autore, per un certo numero di anni prima di (tornare a) essere bene pubblico. In questi casi, la *ratio* economica risiede nella remunerazione della creatività dell'opera d'ingegno[35].

Ma secondo molti studiosi[36], i dati che riguardano la nostra impronta digitale non hanno nulla di creativo e appartengono perciò *direttamente* alla dimensione pubblica, al *public domain*, alla quale noi tutti possiamo accedere per il semplice fatto che quei dati sono stati rivelati.

Questa tesi si scontra tuttavia con il fatto che quei dati «pubblici» non sono sempre rilasciati in un ambiente «pubblico» e, una volta rilasciati, diventano, attraverso una

---

[33] *Ibidem*, p. 61.

[34] BRYNJOLFSSON E. - MITT L. - KIM H., *Strength in Numbers: How does Data-driven Decision Making Affect Firm Performance?*, in *Social science research network paper*, 2011.

[35] BARRETT N., *Will Big Data create a new untouchable business elite?*, in *The Economist*, 2017, p. 13.

[36] MANYIKA J. - CHUI M. - BROWN B., *Big Data: The next frontier for innovation, competition, and productivity*, Mckinsey Global Institute, 2011, pp. 8 ss.

«delega» nel caso di rilascio del consenso, una proprietà privata *de facto* di chi se ne appropri in via esclusiva ai fini dell'attività di profilazione pubblicitaria.

Quindi il punto non è attribuire una proprietà privata su un'informazione che non è prodotto di creatività[37]. Il punto è che il dato inteso come bene pubblico, una volta prodotto, finisce in effetti per essere fatto proprio da coloro che lo usano in via esclusiva a fini di profilazione pubblicitaria, estraendone un valore: un paradosso, difficilmente superabile da visioni, quali quelle avanzate da Josef Drexl, che mettono in discussione, in ragione della sua natura pubblica, l'esistenza di una proprietà privata in capo a coloro che il dato lo producono. Non si comprende, infatti, perché se una proprietà privata viene di fatto *ex post* attribuita sul dato, essa non debba essere riconosciuta inizialmente anche a chi ha generato quella informazione. Tanto più se ciò permette di costruire un mercato trasparente[38].

È vero che attraverso questo «scambio implicito», anche coloro che rilasciano i propri dati ottengono in cambio servizi, ma lo scambio resta, appunto, implicito, non misurato attraverso prezzi dedicati e trasparenti e, dunque, non «internalizzato» dal mercato. Sta tutto qui il nocciolo economico della questione e della novità che abbiamo di fronte.

Nonostante l'operare di regolamentazioni nazionali e internazionali che tutelano la nostra privacy, la natura pubblica dei dati rende sempre più difficile escludere le piattaforme di e-commerce, gli operatori di *cloud computing*, i social network, i motori di ricerca e i servizi correlati, i navigatori, e così via, dalla raccolta e dal consumo dei dati digitali dei cittadini-utenti.

A maggior ragione quando siamo spesso online anche senza il nostro consenso e persino anche senza la nostra presenza. Per esempio, secondo Alien St. John, di «Consumer Reports»[39], quando visitiamo pagine di un sito che contengono le icone di un social network come Facebook e Twitter (che ci permettono di condividere quella notizia sul social di riferimento), quei pixel potrebbero tracciarci anche in assenza di click. Quintarelli, nel *Capitalismo immateriale*[40], cita numerosi esempi di estrazione di dati e immagini che avvengono al di fuori di un rapporto contrattuale e «di consenso» e in modo del tutto inconsapevole da parte di colui che origina il dato.

---

[37] *Ibidem*, p. 12.
[38] *Ibidem*, p. 14.
[39] BARRETT N., *Will Big Data create a new untouchable business elite?*, op. cit., p. 19.
[40] **QUINTARELLI** S., *Capitalismo immateriale. Le tecnologie digitali e il nuovo conflitto sociale*, **Bollati Boringhieri**, Torino, 2019, pp. 65 ss.

In questi passaggi avviene una trasformazione fondamentale del dato (o meglio del suo contenuto informativo) da bene pubblico a bene privato, soltanto che *de facto* la «proprietà privata» del dato non viene riconosciuta in partenza, ma solo in arrivo. Con l'ulteriore paradosso che l'originario generatore del dato non è titolare né di una tutela inibitoria, né di una tutela risarcitoria[41]. Anzi, il «simulacro» della tutela inibitoria, cioè il rilascio del consenso, ove previsto, diventa, paradossalmente, il meccanismo di delega grazie al quale ciò che non era un bene privato (e, secondo alcuni, nemmeno un diritto proprietario alienabile) lo diventa, nella forma dell'*entitlement* su specifici usi, ma esclusivamente per chi lo riceve. È questa la straordinaria novità del dato come bene informazione, rispetto a beni informazione di altra natura, come quelli tipicamente oggetto di proprietà intellettuale[42]. Si capisce allora come non basti affermare che il dato personale deve essere sottratto al mercato in quanto inalienabile (*inalienability rule*)[43], se poi la natura di bene, non rivale e non escludibile, del dato lo trasforma comunque in bene privato e lo consegna, grazie ad uno scambio implicito, al mercato (o meglio a quella parte del mercato che lo utilizza in via esclusiva, prevalentemente all'interno di attività di sfruttamento economico del dato verticalmente integrate, come avviene per molte piattaforme online).

Lo «scambio di dati», al di fuori di una chiara definizione di diritti di proprietà, genera, tecnicamente, una situazione di «fallimento del mercato»: l'incapacità, cioè, del mercato di generare autonomamente un livello ottimale della produzione di un bene e della sua allocazione[44]. È possibile che ciò sia l'inevitabile conseguenza dell'avvento di nuovi modelli di business e che sia persino errato chiamarlo «fallimento del mercato». Ma c'è intanto una grossa novità, per economisti e giuristi, con la quale fare i conti e che non ha trovato finora una ricostruzione unitaria convincente: è il misterioso e inedito rapporto tra diritti e mercato a proposito dei *big data*.

[41] *Ibidem*, p. 71.
[42] *Ibidem*, p. 74.
[43] BARRETT N., *Will Big Data create a new untouchable business elite?*, op. cit., p. 32.
[44] CALABRESI G., *Property Rules, Liability Rules, and Inalienability: One View of the Cathedral*, op. cit., p. 171.

3.1 La «data economy»

Big data, algoritmi e intelligenza artificiale stanno già trasformando moltissimi ambiti: dall'assistenza ospedaliera alla cura sanitaria a distanza, dall'istruzione ai trasporti, dall'informazione giornalistica alla fruizione di contenuti multimediali, dai servizi finanziari e assicurativi alla ricerca di lavoro; dalla *cybersecurity* alla prevenzione di crimini di varia natura; dal controllo del traffico stradale alle previsioni del tempo; dalle attività sportive e di tempo libero all'erogazione e alla fruizione di servizi energetici, dall'agricoltura alla domotica e l'*Internet delle cose* (*Internet of Things*, Iot)[45], e così via. E i campi di applicazione sono moltissimi. Ecco, di seguito, alcuni esempi.

Settore delle comunicazioni

Nel settore delle comunicazioni (telecomunicazioni, fruizione video, informazione in genere).

Il mondo delle comunicazioni si è radicalmente trasformato: escluse poche eccezioni, quelle comunicazioni altro non sono, ormai, che generazione, traffico e scambio di dati. Anche grazie alla rivoluzione delle connessioni 5G è possibile scambiare dati con tempi bassissimi di latenza, interagendo in modo sempre più ricco e complesso con quanto ci circonda. Con lo scambio bidirezionale di dati, l'erogazione di informazione «uno-molti» diventa «molti-molti», secondo il modello di Web e social network: c'è sempre una doppia attenzione, la nostra, verso ciò che riceviamo, e quella di altri, nei confronti dei dati che inviamo (ricordando che la *net neutrality* è il principio che impone di non discriminare gli utenti nell'erogazione di contenuti e di traffico dati)[46]. Tutto ciò, come ha scritto Tim Wu[47], espande il «business del tempo di attenzione» ad ogni forma di comunicazione pubblica o privata, modificando domanda e offerta, nonché vecchie strutture organizzative, costi e modelli di profittabilità. Come ha spiegato Stefano Quintarelli nel suo libro *Capitalismo immateriale*, l'avvento della società digitale comporta che produrre, riprodurre, «archiviare e spedire informazioni

[45] O'LEARY D., *Big Data, Internet of Things and the Internet of Signs*, Intelligent Systems in accounting, finance and management, 2013, pp. 11 ss.
[46] FLEISH E., *What is the Internet of things? An economic perspective*, op. cit., p. 84.
[47] WU T., *I padroni di Internet. L'illusione di un mondo senza confini*, Unwired Media, 2006 , pp. 90 ss.

non costa nulla[48]. Questo ha cambiato le regole del gioco al punto tale che le più grandi compagnie di intermediazione (Facebook, Google, Amazon, Apple, Airbnb, Uber, ma anche molti altri, meno noti al grande pubblico) hanno fatturati che spesso superano quelli di una nazione, con margini da capogiro»[49].

Settore sanitario

Cambia il modo in cui identifichiamo, trattiamo e preveniamo le malattie, migliorando la vita di persone con diverse o ridotte abilità, in maniera non soltanto più efficace ma anche meno costosa per gli individui e per la collettività. Ad esempio, in un'unità specializzata nel trattamento delle nascite premature, presso il *McMaster Children's Hospital* in Canada[50], tecniche basate sui *big data* hanno permesso di monitorare le pulsazioni e il respiro di neonati e, tramite appositi algoritmi, di prevedere possibili infezioni nelle 24 ore precedenti. L'intelligenza artificiale è stata utilizzata per l'analisi di dati radiologici (Mri e XRays), i quali, grazie alla granularità informativa della diagnostica per immagini, permettono di ridurre il numero di operazioni chirurgiche esplorative a fini di diagnosi[51].

Settore bancario e finanziario

L'estrazione ed elaborazione di *big data* aiuta a individuare transazioni fraudolente, a sviluppare analisi delle tendenze commerciali delle imprese, ma anche ad offrire incentivi personalizzati all'uso di carte di credito in funzione dei profili di reddito e di spesa dei clienti.

Secondo un celebre studio[52], la personalizzazione del rapporto tra istituto di credito e cliente è uno degli elementi decisivi per trattenere o attrarre clientela. Ma la vera rivoluzione si chiama *FinTech* e si riferisce sia a vecchi servizi e prodotti finanziari erogati con modalità digitali innovative, sia a nuovi prodotti finanziari, quali ad esempio le criptovalute[53].

---

[48] QUINTARELLI S., *Capitalismo immateriale. Le tecnologie digitali e il nuovo conflitto sociale*, op. cit., p. 88.
[49] *Ibidem*, p. 89.
[50] BRYNJOLFSSON E. - MITT L. - KIM H., *Strength in Numbers: How does Data-driven Decision Making Affect Firm Performance?*, in *Social science research network paper*, 2011, pp. 87 ss.
[51] *Ibidem*, p. 90.
[52] Boston Consulting Group, *Global Retail Banking 2018: The Power of Personalization*, 2018.
[53] *Ibidem*.

Secondo un rapporto di Accenture[54], nel 2017 quasi 30 miliardi di dollari, a livello globale, sono stati investiti in *start-up* innovative dedicate a prodotti e servizi *FinTech*, con una crescita del 16% rispetto all'anno precedente.

Numeri destinati ad aumentare esponenzialmente. Secondo l'indicatore *FinTech Adoption* realizzato da Ernst Young[55], sempre a livello mondiale, un terzo dei consumatori usufruisce già di due o più servizi tecno-finanziari, e circa 1'84% di costoro dichiarano di sapere cosa siano questi servizi e i vantaggi ad essi associati. In questi contesti assume una rilevanza crescente non solo la raccolta di dati ma anche le modalità con le quali i dati vengono utilizzati, registrati, archiviati, resi accessibili.

Un nuovo paradigma tecnologico è rappresentato dalla cosiddetta *blockchain*, una sorta di libro contabile digitale generato da un database organizzato in blocchi (contenenti più transazioni) collegati in nodi e condiviso tra molti partecipanti, che permette di identificare, controllare e approvare ogni transazione - in tempo reale e senza costi aggiuntivi - senza passare da un intermediario, per trasferire informazioni e beni (ad esempio denaro, ma non solo)[56].

Settore del mercato assicurativo

Molte applicazioni utilizzano i dati per aiutare gli assicuratori a fissare premi in maniera più accurata, identificare richieste fraudolente e migliorare le proprie strategie di marketing. Diverse aziende assicuratrici raccolgono dati relativi alla sicurezza della guida, offrendo sconti agli assicurati che si prestano ad un monitoraggio della guida attraverso *app* scaricate su *smartphone* o altri meccanismi volontariamente installati nell'autoveicolo. Questo sistema di «premi» (o *reward*) per i comportamenti virtuosi comporta la raccolta di una grande mole di dati.

Settore della grande distribuzione

Anche questo settore è destinato a cambiare. Da un lato, telecamere digitali vengono abitualmente utilizzate per controllare i livelli delle scorte; dall'altro, le stesse potranno identificare i consumatori attraverso meccanismi di riconoscimento facciale o

[54] Accenture Strategy, *Reworking the Revolution: Are you ready to compete as intelligent technology meets human ingenuity to create the future workforce?,* 2018.
[55] ERNST & YOUNG, *As FinTech becomes the norm, you need to stand out from the crowd*, Global FinTech Adoption Index, 2019.
[56] *Ibidem*.

collegamento con lo *smartphone* degli utenti (ad esempio offrendo loro collegamenti wi-fi gratuiti in negozio), raccogliendo dati sui tempi di permanenza, i prodotti visionati e acquistati, oppure offrendo forme di pagamento e scontistica funzionali ad acquisire informazioni su abitudini di consumo e capacità di spesa. A Seattle e Chicago, ad esempio, si trovano i supermercati «Amazon Go», i primi senza casse all'uscita: gli articoli rimossi dagli scaffali vengono addebitati automaticamente sul conto del cliente una volta che ha varcato l'uscita.

Questo sistema non solo rende più veloce e agevole la spesa per il cliente, ma ne acquisisce dati di consumo permettendo di identificarlo e profilarlo meglio così da fornirgli ulteriori offerte personalizzate.

Settore energetico

Al fine di contrastare l'aumento dei costi (diretti e indiretti, economici e ambientali) di estrazione di petrolio e gas naturale, l'acquisizione e l'elaborazione dei dati degli andamenti di domanda e offerta possono permettere una programmazione più efficiente. Tra i casi citati da Bernard Marr nel suo blog, c'è quello della *Royal Dutch Shell* che ha sviluppato un «giacimento di petrolio basato sui dati»[57] nel tentativo di ridurre i costi della perforazione per la ricerca del petrolio. L'energia elettrica, invece, è un bene non stoccabile e dunque affinché ciascuno di noi possa fruirne occorre un complesso sistema di acquisto programmato, a vari livelli e in vari momenti temporali, che ne assicuri la ridondanza. Un meccanismo che genera anche, inevitabilmente, un certo grado di inefficienza[58]. Oggi, la rivoluzione digitale entra anche nelle nostre case con la domotica e l'Internet delle cose e permette di raccogliere dati sui nostri consumi energetici domestici, misurando e costruendo un profilo adeguato della nostra permanenza e dei nostri bisogni, consentendo agli elettrodomestici di «parlare tra loro» per ottimizzare le decisioni ed evitare sprechi di varia natura[59]. Le «case intelligenti», nelle quali fanno bella mostra assistenti vocali come quelli offerti da Google e Amazon (e che riconoscono e registrano anche la nostra voce come un dato unico e irripetibile di identificazione digitale personale) raccolgono informazioni su di noi e per noi.

[57] MARR B., *Managing and Delivering Performance*, Routledge, 2016, pp. 67 ss.
[58] *Ibidem*, p. 71.
[59] *Ibidem*, p. 75.

Settore del trasporto, della logistica e della consegna postale di pacchi e merci

Le aziende stanno raccogliendo e analizzando i dati relativi ai loro veicoli per migliorare i comportamenti dei conducenti, ottimizzare i percorsi e rendere più efficace la manutenzione dei mezzi. Sul piano dei trasporti, l'utilità della raccolta e dell'elaborazione di dati dispersi è fondamentale per migliorare i servizi, ridurre le varianze relative ai picchi, favorire l'intermodalità. Diventa possibile, per ciascuno di noi, organizzare un viaggio articolato su più mezzi di trasporto, ricevendo proposte personalizzate in termini di prezzo e qualità. Piattaforme come Uber, Flixbus, ecc. nascono dalla capacità di utilizzare i dati e gli algoritmi come strumenti volti a facilitare l'incrocio di domanda e offerta, anche con servizi personalizzati[60].

Ma resta l'auto a guida autonoma, in tutte le sue varianti, l'esempio più emblematico dell'applicazione dei *big data* ai settori industriali tradizionali. La sicurezza della vettura, sotto vari profili, dipende, infatti, dalla capacità del computer di bordo di raccogliere ed elaborare in tempo reale tutti i dati acquisiti dai sensori interni ed esterni. Affinché ciò sia possibile, la mappatura esterna tridimensionale dei percorsi dell'auto, così come i dati sul traffico e il meteo, costituiscono delle precondizioni necessarie. Ma non c'è solo l'automobile. Un servizio sperimentale, avviato nel porto di Livorno, nato da un progetto di collaborazione tra Ericsson, il Porto di Livorno e il Consorzio nazionale interuniversitario delle telecomunicazioni (Cnit), permetterà di elaborare tutti i dati rilevanti all'interno di un sistema complesso di gestione efficiente di partenze e arrivi in uno *smart terminal*, nonché di organizzare la logistica relativa a passeggeri e merci.

Settore agricolo

Anche in questo ambito sono già state avviate modalità di raccolta ed elaborazione dei dati[61]. Gli agricoltori possono accedere ai dati ottenuti dai sensori incorporati nelle loro macchine agricole al lavoro sui campi e anche ai dati aggregati degli altri utenti nel mondo[62]. In Toscana, un progetto nato dalla collaborazione tra Zucchetti ed Ericsson, ha avviato la sperimentazione di un *agrirobot* che gira tra i filari e, grazie a una rete di

[60] REZZANI A., *Big data. Architettura, tecnologie e metodi per l'utilizzo di grandi basi di dati*, op. cit., p. 68.
[61] Ad esempio tramite il sito *myjohndeere.com* negli Stati Uniti.
[62] LUGARD P. - ROACH L., *The era of Big Data and EU/U.S. divergence for refusal to deal*, in *Antitrust*, Vol. 31, 2017, pp. 89 ss.

sensori, analizza lo stato delle uve ed eventuali attacchi di parassiti e insetti, facilita la programmazione del tempo di vendemmia e costruisce una banca dati specifica per ogni pianta. La raccolta di questi elementi permetterà al consumatore di avere informazioni dettagliate sulla qualità del prodotto finale, oltre che sulla sua origine e sulle modalità di coltivazione e trattamento del prodotto primario.

Insomma, guardando ai vari settori e alle varie applicazioni, gli esempi concreti sono moltissimi e hanno tutti in comune quattro elementi: *a)* la raccolta di dati rivelati dai comportamenti di imprese e utenti; *b)* il loro trattamento da parte di appositi algoritmi; *c*) l'elaborazione di modelli predittivi; *d)* la valorizzazione economica dell'attenzione e dei dati in genere. Questi quattro elementi sono gli ingredienti di un nuovo modello di organizzazione capitalistica dei mercati. Nel passaggio storico che viviamo, questi nuovi modelli organizzativi stanno trasformando i mercati tradizionali, registrando, al contempo, l'emersione di grandi e pochi «giganti digitali», spesso denominati *Big Tech*.

La preoccupazione circa la concentrazione di dati, algoritmi, processi e innovazioni nelle mani di pochi colossi a livello globale è oggetto ormai, in tutto il mondo, di dibattito e discussione da parte di parlamenti, governi, autorità garanti preposte alla tutela della concorrenza, della regolazione dei mercati digitali, della privacy e della *cybersecurity*[63].

Si intrecciano qui questioni legate alla competitività e all'affermazione di imprese caratterizzate da dimensioni e da potere di mercato mai esperiti in passato, con temi che riguardano tanto la *cybersecurity* - cioè la sicurezza della connettività in ambiti strategici per ciascun paese - quanto la manipolazione informativa sul Web al fine di influenzare le campagne elettorali e le scelte dei cittadini[64].

Questo dibattito, a livello globale, è oggi animato dalla discussione non soltanto sui rischi e i pericoli derivanti dalla concentrazione di potere economico, ma soprattutto sull'efficacia delle attuali normative antitrust e di regolazione nell'affrontare le nuove dinamiche del capitalismo digitale, garantendo il giusto equilibrio tra libertà d'impresa e tutela della concorrenza da un lato, e tra libertà d'espressione e tutela del pluralismo online dall'altro, mettendo sempre in primo piano il cittadino-consumatore. Al centro

[63] *Ibidem*, p. 93.
[64] *Ibidem*, p. 97.

del dibattito, c'è proprio il tema di come definire regole che garantiscano un accettabile equilibrio tra opposte esigenze.

4. Tipologie di Big Data

A titolo di esempio, citiamo alcune categorie di dati che possono rientrare nella definizione di big data, indicando per ciascuna le caratteristiche importanti[65].

Tabella 1. - Esempi di big data

Caso	Caratteristiche	Esempi di utilizzo
Sensori e DCS	Velocità e volume	Analisi dei guasti e manutenzione predittiva.
Radio Frequency Identification (RFID)[66]	Velocità e volume	Analisi del percorso d'acquisto in un negozio della grande distribuzione. Analisi e tracking delle merci, in combinazione con altri dati (ambientali, geografici, ecc.)
Quotazioni e transazioni su mercati finanziari.	Velocità e volume	Sistemi automatici di high frequency trading; analisi previsionale.
Dati da strumenti	Velocità e volume	Riconoscimento di pattern. Simulazioni.
Dati astronomici	Volume e varietà	Analisi del quantitativo enorme di dati raccolti da osservatori e radiotelescopi.
Dati metereologici	Volume	Previsioni meteo. Monitoraggio di eventi atmosferici estremi.
Informazioni sanitarie	Volume e varietà (diversità di formati)	Ministero della salute, enti di ricerca: identificazione e monitoraggio della diffusione di malattie.
Dati fiscali, bancari e patrimoniali	Volume	Il Ministero delle Finanze, Agenzia delle Entrate e Guardia di Finanza possono utilizzare le enormi banche dati a loro disposizione per l'identificazione di comportamenti anomali che indicherebbero casi di evasione fiscale.
Social Network	Varietà: diversità di formati, dati semi-strutturati	Sentiment Analysis (come si sta parlando della nostra azienda? Come è stato accolto il nuovo prodotto?) CRM (Customer Relationship Management) Utilizzo da parte di servizi di intelligence.

[65] BOURREAU M. - DE STREEL A. - GRAEF I., *Big Data and competition policy: market power, personalised pricing and advertising*, in *Cerre project report*, 2017, pp. 76 ss.

[66] RFID è una tecnologia per l'identificazione automatica di oggetti, animali o persone basata su dispositivi elettronici.

Blog, Forum	Varietà: Dati semi-strutturati	Sentiment analysis Utilizzo da parte di servizi di intelligence.
Web server log	Volume	Analisi del traffico sui web server, identificazione dei comportamenti di navigazione degli utenti.
Log del traffico di un	Volume e Velocità	Utilizzo da parte di provider.
Dati provenienti da sistemi di sorveglianza	Volume, Velocità e diversità di formati	Utilizzo da parte di polizia, enti di vigilanza, servizi di intelligence.
Documenti	Volume e assenza di struttura	Fraud detection: per esempio, l'analisi di richieste di risarcimento effettuate alle assicurazioni possono essere analizzate e associate a casi fraudolenti ed esaminate con più attenzione.
Dati geografici	Volume, velocità	I dati provenienti da sistemi GIS possono essere utilizzati assieme ad altri dati per scopi diversi.

Alle categorie esemplificate della Tabella 1 aggiungiamo i *dati dei sistemi operazionali* che in certe aziende raggiungono volumi ragguardevoli. Anche in Italia alcuni grandi gruppi bancari si stanno dotando di sistemi big data per l'analisi delle transazioni sui conti correnti e sulle carte di credito, utilizzando tali dati al loro massimo livello di dettaglio e con profondità storiche che non è possibile raggiungere con i mezzi tradizionali (se non a costi poco giustificabili)[67].

5. I dati aziendali

In azienda possiamo considerare i dati al pari di qualsiasi altro *asset*[68]. I dati e le tecnologie di analisi, soprattutto nell'ultimo decennio, hanno assunto il carattere di risorsa indispensabile per la gestione aziendale. In questo paragrafo proponiamo una panoramica utile a comprendere il patrimonio dei dati e, nell'analizzarlo, consideriamo diversi punti di vista: le fonti, i supporti tecnologici e il tipo di struttura.

[67] COLANGELO G. - MAGGIOLINO M., *Big Data as a misleading facility*, in *European Competition Journal*, 2017, p. 32.
[68] *Ibidem*, p. 39.

5.1 Le fonti

Per quanto riguarda la provenienza dei dati possiamo operare una prima generale distinzione tra fonti interne e fonti esterne all'azienda[69].

Fonti interne

Individuiamo più tipologie di fonti interne che sono in parte sovrapponibili: le fonti operazionali, i *data warehouse* o i *data mart*.

Fonti operazionali

Un primo gruppo è formato dalle fonti operazionali, cioè quelle che fanno riferimento all'attività operativa giornaliera dell'azienda e che, per questo motivo, variano a seconda della tipologia di business e settore economico. Per un'azienda industriale, alcuni esempi di fonti operazionali possono essere i seguenti[70]:

- Applicativi di gestione della produzione. Essi registrano le quantità di materie prime utilizzate, i servizi e i beni accessori consumati (elettricità, acqua, combustibili), le quantità prodotte, ecc.
- Applicativi di gestione degli acquisti. I sistemi a supporto di quest'area si occupano di registrare ogni ordine di acquisto e i movimenti di magazzino.
- Applicativi di gestione degli ordini e delle consegne. Sono sistemi per la registrazione di ordini ricevuti dai clienti, attività di consegna e movimentazione del magazzino prodotti.
- Applicativi di contabilità. A fronte di acquisti e di vendite occorre registrare i movimenti contabili che scaturiscono dall'emissione di fatture di vendita o dal ricevimento di fatture d'acquisto, oltre a ogni altra movimentazione di cassa o banca.
- Applicativi di gestione del personale. Anche le risorse umane prevedono una contabilità che riguarda soprattutto la gestione degli stipendi, alla quale però si possono aggiungere attività quali la gestione di obiettivi, premi, malattie, infortuni, ecc.
- Applicativi di gestione del cliente. I cosiddetti applicativi di CRM (*Customer*

[69] ROCCASALVA G., *I Big Data e gli strumenti di visualizzazione analitica: interazioni e studi induttivi per le P.A.*, Apogeo, Milano, 2018, pp. 65 ss.

[70] *Ibidem*, p. 72.

Relationship Management) consentono la gestione completa del cliente, dalla registrazione dei dati anagrafici, fino alla gestione delle campagne di marketing.

Per un'azienda bancaria, le fonti operazionali sono in parte simili a quelle dell'esempio precedente (contabilità, gestione del personale, CRM), ma a esse si aggiungono applicativi tipici dell'attività bancaria: la gestione dello sportello, il *back office*, le applicazioni di gestione e valutazione degli strumenti finanziari, gli applicativi per la valutazione del rischio e l'erogazione di finanziamenti, ecc.

Le aziende commerciali, come quelle della grande distribuzione, presentano oltre ai sistemi di contabilità, gestione acquisti, magazzino e gestione del personale anche i sistemi di rilevazione delle vendite alle casse ed emissione degli scontrini fiscali oppure gli applicativi per la gestione delle promozioni e delle tessere clienti.

I dati operazionali per alcune aziende possono assumere volumi rilevanti. Basti pensare all'ambito bancario, o all'ambito industriale ove vi possono essere
sistemi, legati alla produzione, che generano enormi quantità dati.

Si tratta, di solito, di DCS (*Distributed Control Systems*)[71], cioè di sistemi computerizzati utilizzati per il controllo di impianti industriali. Gli elementi controllanti non sono centralizzati, ma sono distribuiti sull'impianto. I componenti del sistema, connessi tramite una rete che consente il controllo, la comunicazione e il monitoraggio, generano dati relativi allo stato degli impianti mediante sensori legati al componente stesso. Le rilevazioni dei dati possono avvenire ad intervalli temporali molto piccoli e ciò, assieme alla presenza anche di migliaia di sensori, porta a produrre una mole elevata di valori.

Data warehouse e data mart

L'analisi effettuata direttamente sui sistemi operazionali è sconsigliabile per diverse motivazioni. Nella migliore delle ipotesi, ciascuno degli applicativi appena descritti è semplicemente un modulo di un software ERP (*Enterprise Resource Planning*), acquistato "chiavi in mano" da un produttore come SAP, Microsoft oppure Oracle. Purtroppo, nella maggioranza dei casi, accade che non vi sia un'unica applicazione che

[71] DE MAURO A., *Big Data Analytics. Analizzare e interpretare dati con il machine learning*, op. cit., p. 89.

gestisce le attività del livello operativo. Ci si trova spesso di fronte ad una pletora di software, ognuno dei quali è basato su tecnologie differenti (database relazionali, basi dati non relazionali) e su prodotti di *vendor* diversi. In presenza di numerose applicazioni, l'uniformità e la coerenza dei dati non sono per nulla garantite, poiché ci si trova in una situazione in cui dati fondamentali, quali le anagrafiche (clienti, fornitori, prodotti, piano dei conti, ...), sono replicati e manipolati in ciascuno dei software, presentando differenze di formati, di completezza o di aggiornamento dei dati[72].

Un altro aspetto da tenere in considerazione riguarda il disegno delle basi dati sottostanti agli applicativi operazionali. Esse sono di tipo OLTP (*On Line Transaction Processing*) e il loro modello dati è fortemente normalizzato per favorire non tanto le letture e l'analisi di grandi quantità di record, quanto le attività transazionali (inserimenti, cancellazioni, modifiche dei dati). La normalizzazione, se da un lato favorisce l'attività transazionale, dall'altro incrementa notevolmente il numero di tabelle utilizzate per contenere i dati. Per ricostruire un formato tabellare denormalizzato, più adatto ad essere letto da un utente finale, occorreranno diverse operazioni di JOIN[73], che complicano l'attività di estrazione dei dati. Inoltre, occorre considerare che solitamente i sistemi operazionali offrono una limitata profondità storica dei dati e, a volte, essa manca del tutto. Molto spesso, anche in presenza di dati storici, ricostruire la situazione dei dati nel passato è un'attività complessa.

L'insieme dei fattori descritti rende piuttosto difficoltosa l'analisi dei dati effettuata direttamente sulle fonti operazionali. La strada più corretta consiste nella creazione di un *data warehouse* o di una serie di *data mart*, cioè di database che contengono dati integrati, coerenti e certificati afferenti a tutti i processi di business dell'azienda (o per lo meno alla maggior parte di essi) e che costituisce il punto di partenza per le attività analitiche del sistema di *Business Intelligence* (BI). La *Business Intelligence* è un sistema di modelli, metodi, processi, persone e strumenti che rendono possibile la raccolta regolare ed organizzata del patrimonio di dati generato da un'azienda. Inoltre, attraverso elaborazioni, analisi o aggregazioni, ne permette la trasformazione in informazioni, la loro conservazione, reperibilità e presentazione in una forma

[72] *Ibidem*, p. 90.
[73] Nel linguaggio SQL, la JOIN consente di mettere assieme i dati di due tabelle.

semplice, flessibile ed efficace, tale da costituire un supporto alle decisioni[74] strategiche, tattiche ed operative.

In alcuni casi accade che il *data warehouse* sia quasi esattamente sovrapponibile alle fonti dati operazionali (come contenuti, ma non come disegno!). In esso quindi confluiscono tutte (o quasi) le sorgenti dei dati aziendali, senza che i dati subiscano filtri o aggregazioni. In altri casi, però ci si scontra con vincoli tecnologici o di costo che impongono versioni ridotte del *data warehouse*. La riduzione può avvenire sia in termini di profondità storica conservata all'interno della base dati analitica, sia in termini di allargamento della granularità dei dati (per esempio, dati settimanali e non giornalieri).

Basi dati ad hoc

Definiamo *basi dati ad hoc*, quelle create per specifiche esigenze analitiche e che contengono particolari rielaborazioni di dati operazionali o di dati presenti nel *data warehouse*. Non è raro trovare in azienda database di "proprietà" del singolo analista, che contengono al loro interno dati molto preziosi, soprattutto se fossero condivisi con altri analisti o funzioni aziendali.

Fonti esterne

Non è raro che, per particolari analisi, si renda necessario l'utilizzo di dati provenienti dall'esterno, in aggiunta ai dati generati internamente all'azienda. In taluni casi potrebbe trattarsi semplicemente di integrazioni alle anagrafiche (per esempio l'anagrafica ISTAT dei comuni italiani), mentre in altri casi i dati esterni sono il vero e proprio oggetto delle analisi. Come esempio citiamo l'analisi del *sentiment*, volta a verificare quale sia l'opinione delle persone che scrivono sui social rispetto ad una certa tematica, un certo prodotto o una certa azienda. Per realizzare questo tipo di attività occorrono dati provenienti dai social network (Facebook, Twitter,...), dai blog o da forum e dunque esterni all'azienda.

Il reperimento e l'utilizzo di dati esterni pongono alcuni problemi. Uno di essi, forse il principale, consiste nella loro qualità, che potrebbe presentare difetti di accuratezza,

[74] I sistemi di Business Intelligence sono anche definiti come *Decision Suppott System* (DSS).

completezza e coerenza[75]. Occorre precisare che la problematica relativa alla qualità riguarda anche i dati interni; tuttavia, sui dati esterni l'azienda non ha alcuna possibilità di manovra e spesso è difficile anche la semplice verifica del livello qualitativo.

5.2 Tipi di supporto

I dati interni ed esterni sono presenti su supporti diversi. Le fonti operazionali sono ospitate, salvo poche eccezioni, in database relazionali[76]. Può variare la tecnologia o il *vendor*, ma tali supporti sono facilmente accessibili tramite interrogazioni con il linguaggio SQL.

I casi più complessi sono quelli legati all'uso di tecnologie legacy (*mainframe*), che si riscontrano più spesso in ambito bancario. Nel caso del mainframe non è raro che i dati siano esposti tramite la produzione di file di testo in vari formati (CSV, TSV, record a lunghezza fissa).

I *data warehouse* si trovano esclusivamente su RDBMS, sia con tecnologia SMP sia con tecnologia MPP. I sistemi comunemente utilizzati per ospitare RDBMS rientrano

---

[75] La qualità è determinata da un insieme di caratteristiche, che elenchiamo di seguito:

- *Completezza*: presenza di tutti i dati necessari per descrivere un'entità, una transazione o un evento. Per esempio la mancanza di valori nei campi anagrafici di un'entità di business rende quell'anagrafica incompleta.
- *Consistenza*: assenza di contraddizioni nei dati. La consistenza è legata al concetto di equivalenza dei dati. Per esempio, in una banca, la somma del saldo di fine mese precedente di un conto corrente con i movimenti attivi e passivi, deve essere uguale al saldo di fine mese corrente; ciò significa che le rilevazioni dei saldi e dei movimenti devono essere tra loro consistenti. La consistenza è anche dettata dalle regole di business, che differiscono da azienda ad azienda oppure da settore a settore.
- *Accuratezza*: riguarda la conformità ai valori reali, cioè la correttezza dei valori stessi. Il punto di partenza per l'accuratezza dei dati risiede nella loro integrità.
- *Assenza di duplicazione*: campi, record o tabelle devono essere presenti soltanto una volta, evitando duplicazioni nello stesso sistema oppure in sistemi diversi. Oltre alla necessità di una doppia manutenzione, la duplicazione dei dati incide negativamente sulla qualità, poiché è possibile una mancata sincronizzazione tra le copie dei dati.
- *Integrità*: si utilizza il termine integrità con riferimento ai database relazionali. Essi infatti garantiscono - attraverso strumenti quali i tipi di dato, i check constraint, le chiavi primarie e le chiavi esterne - che i dati rispettino alcuni vincoli: per esempio che in una colonna vi siano soltanto dati dello stesso tipo (numerici, stringa, data, ecc.); oppure che non esistano due righe uguali all'interno di una tabella; oppure, ancora, che in caso di relazione tra due tabelle, una colonna, sui cui è definita tale relazione, dovrà contenere solo valori appartenenti ad una colonna di un'altra tabella.

[76] I database relazionali sono costituiti da tabelle (dette anche *relazioni}* e sono realizzati secondo i principi della teoria relazionale. Quest'ultima, creata da E.E Codd nel 1970, utilizza un insieme di termini matematici per definire concetti che, comunemente, sono noti con nomi facenti parte della terminologia SQL. Per esempio, il termine *relazione* indica la tabella, la *tupla* indica la riga e l'*attributo* indica la colonna. Nella teoria relazionale, la relazione è costituita da un insieme di tuple aventi gli stessi attributi. All'interno di un attributo (o colonna) i dati hanno lo stesso dominio e sono soggetti agli stessi vincoli. L'accesso ai dati avviene tramite query che possono ritornare tuple (select), opportunamente filtrate secondo predicati logici, combinare tabelle con operazioni di join, oppure modificare i dati con operazioni di insert, update o delete.

nella tipologia degli SMP, Symmetric MultiProcessing: essi sono costituiti da più processori che condividono lo stesso sistema operativo, la stessa memoria RAM e lo stesso *bus* di Input/Output[77]; per questo sono detti *shared everything*. I sistemi SMP sono molto efficienti nelle applicazioni OLTP, ma presentano limiti quando li si utilizza per elaborare grandi volumi di dati. Il limite è dato dal sovraccarico del bus di sistema che costituisce un inevitabile collo di bottiglia. I sistemi MPP, Massive Parallel Processing, si differenziano dagli SMP per il fatto che ogni processore utilizza risorse ad esso dedicate (*shared nothing*), sia per quanto riguarda la RAM sia per quanto riguarda il bus di I/O. I processori comunicano tra di loro attraverso un'interfaccia di *messaging*.

Le limitazioni dovute alla condivisione del bus vengono meno rendendo così le architetture MPP adatte alla gestione di grandi quantità di dati.

Le basi dati ad hoc potrebbero essere ospitate sia da database reazionali, sia da fogli di calcolo. In quest'ultimo caso la loro lettura potrebbe risultare problematica.

5.3 I tipi di struttura

Un altro punto di vista rispetto al quale possiamo classificare i dati riguarda la loro struttura, o meglio, la presenza o l'assenza di una struttura che consenta di identificare agevolmente ogni attributo. Distinguiamo quindi le seguenti tipologie[78]:

- Dati strutturati. Si tratta dei dati che sono rappresentabili in formato tabellare all'interno di un database relazionale, oppure tramite un formato quali l'XML o il JSON, che assieme ai dati contengono i metadati che definiscono i nomi dei campi e la loro struttura. "Rappresentabili in formato tabellare" significa che i dati non si devono per forza trovare in un database, ma potrebbero trovarsi, per esempio, in un file di testo in formato CSV, nel quale i singoli campi sono ben separati e identificabili (quindi il file potrebbe essere importato senza problemi in una tabella).
- Dati non strutturati. Non strutturati sono i dati che non ha senso rappresentare in formato tabellare, anche se essi potrebbero essere inseriti in un database

[77] II *bus* è il canale che collega il processore a tutti i dispositivi (per esempio, gli hard disk) e attraverso cui passano i dati.

[78] REZZANI A., *Big Data Analytics. Il manuale del data scientist*, Apogeo Education, Milano, 2017, pp. 74 ss.

relazionale: non avendo una struttura, infatti, l'intero blocco di dati sarebbe contenuto in un unico campo all'interno di una tabella, rendendo tale struttura completamente inutile. Alcuni esempi di dati non strutturati sono:

  - Il contenuto testuale di un documento (email, pdf, *tweet*, post di un blog).
  - I byte di un'immagine, di un video, di un file sonoro.

- Dati semi-strutturati. I dati semi-strutturati presentano una parte dotata di struttura e una parte non strutturata. Per esempio un documento Word, o PDF, possiede una serie di metadati che sono molto ben strutturati (titolo, autore e molto altro), mentre il corpo del documento è costituito da testo. Lo stesso vale per le immagini che all'interno del file presentano una serie di metadati che descrivono lo scatto, le impostazioni della fotocamera, la data e ora e addirittura le coordinate GPS.

Evidenziamo il fatto che per lavorare con i dati non strutturati occorre effettuare su di essi trasformazioni in grado di fornire un risultato che possegga una struttura. Citiamo di nuovo come esempio l'analisi del *sentiment*, realizzata sul contenuto di una serie di tweet. A parte alcune operazioni preliminari con le quali si cerca di "pulire" il testo dei *tweet*, l'operazione fondamentale da compiere per procedere all'analisi con sistemi predittivi (dato un *tweet* vogliamo capirne il *sentiment* automaticamente), consiste nel creare la cosiddetta *document-term matrix*, cioè una matrice che ha sulle righe i singoli *tweet* (i documenti) e sulle colonne le singole parole (*term*)[79]. In corrispondenza dell'incrocio tra documento e parola vi sarà un numero che indica le occorrenze di quella parola in quello specifico documento. La costruzione della *document-term matrix* è dunque la creazione di un *dataset* strutturato a partire dall'insieme di dati non strutturati.

5.4 La provenienza

Un'ulteriore classificazione distingue tra dati generati dalle persone e dati generati dalle macchine. Alla prima categoria corrispondono sia dati esterni, come quelli provenienti dai *social network* o dai siti di *e-commerce*, sia alcuni dati interni imputati

[79] *Ibidem*, p. 76.

manualmente (operazioni di *data entry*)[80]. I dati generati dalle macchine comprendono numerose casistiche: dai dati provenienti da strumenti di misurazione, ai valori prodotti dai sensori, dai dati di *log* di *web server* e *router* ai sistemi di calcolo che da dati grezzi producono automaticamente nuovi valori e, ancora, dai *device* di lettura di codici a barre (per esempio, di un supermarket) ai sistemi analoghi utilizzati dalle aziende di logistica. Il dato generato automaticamente è di solito meno soggetto a problemi di qualità, anche se malfunzionamenti del dispositivo che genera i dati possono creare numerose anomalie (pensiamo ad un sensore di temperatura che si guasta)[81].

6. Attori aziendali e dati

Gli attori aziendali presentano esigenze e requisiti diversi circa le informazioni che consentono loro di svolgere il proprio compito.

6.1 I manager

Il management si occupa di attività strategiche e tattiche. Le prime consistono nella definizione degli obiettivi aziendali e delle politiche volte al loro raggiungimento. Le attività tattiche riguardano invece l'allocazione efficace ed efficiente delle risorse al fine di conseguire gli obiettivi e l'attività di controllo sul raggiungimento degli stessi (per esempio budget, programmazione della produzione). I manager hanno quindi bisogno dei dati che risiedono in due grandi categorie di applicativi[82]:

- *Knowledge Management System* (KMS). Si tratta dell'insieme degli strumenti software per la ricerca, l'identificazione, la strutturazione di tutte le informazioni riguardanti le attività svolte in azienda. Un esempio è individuabile nei software di gestione documentale, che consentono di creare, organizzare e consultare vere e proprie biblioteche elettroniche costituite dai documenti aziendali.
- Sistemi di *Business Intelligence* (BI), che sono formati da un insieme di strumenti il cui compito finale è quello di fornire un supporto alle decisioni, tramite la trasformazione dei dati aziendali in informazioni. Il punto centrale dei sistemi di business intelligence è il *data warehouse*, che fornisce dati agli

[80] HARFORD T., *Big data are we making a big mistake?*, in *Financial Times*, 28 marzo 2014, p. 5.
[81] *Ibidem*, p. 8.
[82] O'LEARY D., *Big Data, Internet of Things and the Internet of Signs*, op. cit., p. 75.

strumenti di reportistica e ai processi di *predictive analytics*.

Il management richiede informazioni analitiche di sintesi, presentate anche con l'ausilio di un supporto grafico tale da offrire, attraverso un semplice sguardo, la percezione dell'andamento dell'azienda o di singoli settori aziendali. Spesso per arrivare a produrre un output adatto al management occorrono molte fasi di elaborazione, che partono dai dati grezzi e arrivano alla determinazione dei *Key Performance Indicator* (KPI) aziendali[83].

Le esigenze del management non si fermano però alla descrizione del passato. È sempre maggiore il desiderio di anticipare gli eventi ponendo in essere un'attività predittiva volta a raggiungere un vantaggio competitivo nel mercato.

Gli strumenti per realizzare analisi di questo genere non mancano e la crescente disponibilità di dati ne aumenta l'efficacia e offre nuovi spunti analitici. I dati di cui parliamo appartengono sia alla categoria dei dati esterni, sia ai dati interni che con le tecnologie big data sono sfruttabili al livello di massimo dettaglio e con una profondità storica molto elevata. Le analisi predittive sono comunque destinate a produrre risultati di sintesi, anche se originano da dati di estremo dettaglio.

6.2 II personale esecutivo

Il personale esecutivo si occupa dell'operatività corrente (gestione ordini, magazzino, fatturazione,...) che permette all'azienda di funzionare e difficilmente richiede dati sintetici per svolgere i propri compiti: la natura delle proprie attività esige invece dati di dettaglio, forniti con tempestività. Tali dati sono contenuti nei seguenti sistemi applicativi[84]:

- *Enterprise Resource Planning* (ERP). Sono strumenti gestionali che interessano praticamente tutte le aree aziendali, dalla gestione degli ordini alla fatturazione, dalle paghe al bilancio, dai pagamenti alla contabilità.
- *Supply Chain Management* (SCM). I software di questa categoria gestiscono la catena di fornitura (sia relativa agli acquisti, sia relativa alle vendite), velocizzando la trasmissione di ordini di acquisto, ottimizzando le scorte di magazzino, definendo piani di produzione in base alla domanda, monitorando i

[83] BARRETT N., *Will Big Data create a new untouchable business elite?*, op. cit., p. 54.
[84] MARR B., *Managing and Delivering Performance*, op. cit., p. 98.

processi di consegna.

- Software di *Customer Relationship Management* (CRM). Supportano l'attività di contatto con il cliente e forniscono strumenti per la gestione delle vendite, delle campagne di marketing e del *customer service*.
- Sistemi di Business Intelligence. Sono in grado di supportare l'attività operativa con analisi di dettaglio fornite in tempo quasi reale; si parla in questo caso di *Operational Business Intelligence.*

Le prime tre tipologie di software sono generalmente supportate da database relazionali che, a loro volta, rappresentano le fonti per alimentare il *data warehouse* e il sistema di Business Intelligence[85].

Il personale operativo ha esigenze analitiche che possono riguardare aspetti ben più granulari rispetto alle tematiche trattate dai manager e, come questi ultimi, sono beneficiari di nuove opportunità che si generano con le tecnologie big data: più dati e tecniche predittive che consentono di ottimizzare i processi di produzione, vendita, CRM, ecc.

6.3 I *data scientist*

Abbiamo già nominato la figura del *data scientist*, ovvero colui che in azienda ha il compito di svolgere l'analisi dei dati[86]: la sua attività è in particolar modo attinente a quei processi che richiedono l'utilizzo di tecniche avanzate, informatiche, statistiche o di machine learning.

Proprio per quanto detto, il *data scientist* diventa una figura centrale in azienda: egli porta con sé un insieme di competenze che consentono di utilizzare i dati per generare vantaggi competitivi.

È innegabile che anche il termine *data scientist* sia diventato una *buzzword*, al pari di quanto è successo per "big data", ma al di là delle mode del momento, l'esigenza di un profilo con forti competenze analitiche è ormai una necessità per quasi tutte le aziende.

Il profilo ideale di un *data scientist* comprende numerose competenze, che è molto difficile trovare in un'unica persona[87]:

[85] *Ibidem*, p. 101.
[86] REZZANI A., *Big Data Analytics. Il manuale del data scientist*, op. cit., p. 121.
[87] Ibidem, p. 124.

- Competenze informatiche riguardanti in particolare:
    - Database relazionali e linguaggio SQL
    - Hadoop e il suo "ecosistema" di software
    - Spark
    - Database NoSQL, in particolare quelli presenti nella piattaforma Hadoop
    - Linguaggi di programmazione fortemente orientati all'analisi dei dati, quali R o Python
    - Tool di analisi e visualizzazione avanzati.
- Competenze statistiche.
- Conoscenza delle tecniche di predictive analytics e machine learning.
- Conoscenza dei processi di gestione della qualità del dato.
- Conoscenza dei processi di business e dell'organizzazione aziendale.
- Conoscenza delle problematiche e delle sfide del dominio settoriale di interesse (industria, finanza, telecomunicazioni, ...).
- Capacità di comunicare i risultati dell'analisi a tutti i destinatari.
- Capacità di colloquiare con i manager e di supportarli nelle decisioni critiche.

È ovvio, quindi, che il singolo *data scientist* avrà livelli elevati di competenza solo in alcuni di questi campi e potrà, attraverso l'esperienza all'interno dell'azienda, acquisire conoscenze ulteriori, in particolare legate al business e ai processi interni.
Un approccio che funziona quasi sempre molto bene consiste nella creazione di un team di *data scientist*, ognuno dei quali apporta al gruppo competenze diverse e volte verso i differenti aspetti informatici, statistici o di business.

7. La catena del valore del dato

Volume, varietà, velocità dei dati ne generano il valore.
Il *volume* dei dati, come abbiamo già visto, rappresenta sicuramente la caratteristica che più facilmente si può accostare ai *big data*. L'unità di misura più idonea appare quella dei *zettabyte*: uno *zettabyte* corrisponde a una capacità di archiviazione pari a oltre 36.000 anni (in termini di durata) di video in Hd ovvero una pila composta da 250 miliardi di Dvd.
Secondo Icd (*International Data Corporation*), si prevede, nel 2025, a livello globale,

una massa di dati di ammontare pari a 163 *zettabyte,* con una crescita del volume di circa dieci volte rispetto a quello registrato nel 2016[88]. In particolare, una mole di dati sempre maggiore deriverà dal consumo di video online e dalla presenza di sensori legati all'Internet delle cose - dalla domotica alle automobili - dei quali ci si aspetta, già nel prossimo quinquennio, una crescita esponenziale, grazie all'avvento delle connessioni 5G. Siamo entrati nell'«età dello *zettabyte*»[89] e ciò sperimenta un nuovo rapporto tra l'uomo e la società dell'Ict, nell'archiviazione, valorizzazione e memorizzazione di tutte queste informazioni.

C'è poi la *varietà* dei dati che si riferisce all'eterogeneità delle fonti sorgenti dei dati, dei formati con cui vengono acquisite le informazioni (tradizionali/strutturate e, soprattutto, non strutturate) e della rappresentazione e dell'analisi (anche semantica) dei dati immagazzinati.

Sebbene i dati *strutturati* cioè organizzati secondo una precisa struttura, siano spesso quelli che contengono una densità di informazioni maggiore, circa 1'80% dei dati oggi disponibili ha una natura *non strutturata* cui è associato un enorme potenziale informativo e semantico. In questi casi, occorrono tecniche molto sofisticate per trattare dati così diversi e tramutarli in informazione (immagini, foto, testi, email, Rss *feed*, video, sensori, *social media*, ecc.)[90]. Ci sono poi i dati *semi-strutturati*. Un caso tipico è quello delle *e-mail*: qualsiasi servizio di posta elettronica presenta una serie di dati strutturati (che può essere raccolto e organizzato in *database relazionali*). Il corpo dell'e-mail, tuttavia, è generalmente composto da un testo non strutturato e dati con formati assai diversi: immagini, video, audio, ecc. Per analizzare il comportamento degli utenti sul *web* e costruirne una profilazione sono disponibili diversi strumenti che tracciano e memorizzano: *cache* del computer, *cookies,* cronologia dei siti visitati e vari strumenti di *advertising online* personalizzato, dai banner ai *pop-up*, agli spot realizzati in base alle preferenze espresse dagli utenti nella loro navigazione[91].

Infine, la *velocità* dei dati risulta connessa, in primo luogo, alle tempistiche con cui le banche dati vengono alimentate, in particolare all'alta frequenza con cui i dati

---

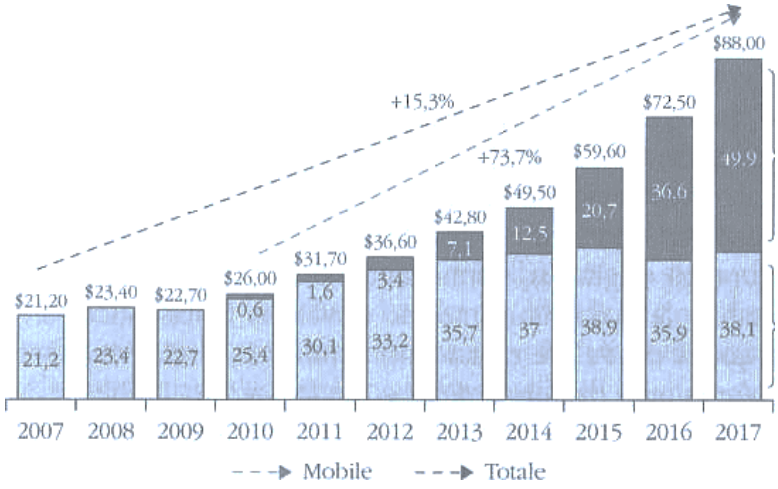
[88] QUINTARELLI S., *Capitalismo immateriale. Le tecnologie digitali e il nuovo conflitto sociale*, op. cit., p. 153.
[89] *Ibidem*, p. 154.
[90] ROCCASALVA G., *I Big Data e gli strumenti di visualizzazione analitica: interazioni e studi induttivi per le P.A.*, op. cit., p. 87.
[91] *Ibidem*, p. 89.

circolano da un punto di origine a uno di raccolta. La velocità non riguarda esclusivamente il flusso di dati, ma anche la necessità di processare i dati in maniera rapida, per prendere decisioni ad un ritmo sempre più veloce, spesso in tempo reale (cd. *real-time action* e *real-time processing*)[92].

C'è chi si è divertito a contare oltre 70 *v* dei dati, includendo ad esempio la *veridicità* (la fiducia che in essi si può riporre), la *valenza* (cresce nel tempo e riguarda le connessioni fra dati), la *visualizzazione* dei dati (il modo in cui riusciamo a rappresentarli)[93].

La *v* più importante, tuttavia, è quella che deriva da tutte le altre ed è legata alla capacità di estrarre *valore* dai *big data*. Ovviamente l'aspetto preponderante risiede nell'attività di raccolta pubblicitaria online (*digital advertising*) per la commercializzazione di prodotti e servizi indirizzata verso una domanda già profilata (Figura 1).

FIG. 1. - Andamento dei ricavi pubblicitari online nel mondo (2007-2017).



Fonte: Elaborazioni Agcom (2018) su dati lab.

Accanto al valore privato per le imprese (e per i consumatori che risparmiano costi transattivi), c'è anche il valore pubblico dei dati che possono essere impiegati per il disegno di politiche volte ad accrescere il benessere complessivo della società.

[92] REZZANI A., *Big data. Architettura, tecnologie e metodi per l'utilizzo di grandi basi di dati*, op. cit., p. 111.
[93] *Ibidem*, p. 112.

Affinché i dati acquisiscano davvero un valore economico essi devono essere funzionali allo svolgimento di analisi economico-statistiche.

Secondo uno studio condotto nel 2013, e pubblicato dalla «*Mit Technology Review*»[94], soltanto lo 0,5% dei dati disponibili sarebbe oggetto di analisi. Lo sfruttamento dei dati digitali generati dagli utenti, inoltre, permette di rispondere a specifiche domande di ricerca (uso primario), quando queste esistono, e di sfruttarne, nel tempo, il loro «valore opzionale» (uso secondario), di cui quasi sempre non si conosce neppure l'esistenza, al momento della raccolta dei dati. Il riutilizzo dei dati è alla base dei numerosi progetti che Google e altre società della rete hanno in cantiere (e sono resi al pubblico spesso in versioni c.d. *beta*, ossia sperimentali)[95].

Volendo individuare una «catena del valore del dato», il primo scalino è dato dalla sua *acquisizione*. Essa dipende anche dal tipo di dato raccolto. Oggi si raccolgono dati da una varietà di fonti molto eterogenee. Per evitare il rischio che un sistema così diffuso di raccolta possa generare un fenomeno di *over-colleclion*, cioè di ridondanza del dato, si opera una seconda fase di lavorazione che riguarda la *preparazione* e la *conservazione* del dato per gli usi successivi (*data silos*, *data warehouse*, *data marts*)[96].

Oggi, il nuovo paradigma del *data lake* si basa sulla condivisione di dati grezzi - strutturati, semi-strutturati e destrutturati - nel loro formato originario, permettendone così l'analisi e, in ultima istanza, la massima estrazione di valore. Le fasi successive della catena del valore riguardano l*'analisi* e le relative attività di *esplorazione*, *trasformazione* e *modellazione*. Infine, si ha l*'immagazzinamento*, un processo che deve rispettare precisi criteri che consentano una facile scalabilità e memorizzazione. Per grandi aziende o per grandi quantità di dati, la realtà odierna è quella dei *database distribuiti*. Nel mondo dei *big data*, i dati possono essere distribuiti sulle memorie di massa dei diversi computer (o nodi) che costituiscono la rete di un'organizzazione i cui nodi possono anche essere fisicamente molto distanti. Lo stadio finale del dato è quello in cui da semplice informazione si trasforma in *conoscenza e visione* (o *wisdom*). Tutte queste fasi della *catena del valore del dato* consentono così di modellare il sistema dei *big data* e, di conseguenza, di identificare i vari passaggi attraverso

[94] OVI A., JACOBELLI G.P., *Lo sfruttamento dei dati digitali*, in *MIT Technology Review*, 2013.
[95] *Ibidem*.
[96] *Ibidem*.

cui generare valore, e, più in generale, conoscenza. Ma quanto vale il nostro dato?[97]
L'analogia dei dati con il petrolio funziona solo in parte: un barile di petrolio ha di per sé un valore, chiaro e osservabile, ma ciò non vale per i dati. Mentre il prezzo di equilibrio sul mercato del petrolio deriva dalla sua scarsità, nella tensione tra domanda e offerta, ciò non vale per il prezzo di mercato dei dati che, al contrario del petrolio, possono essere riprodotti, riutilizzati e riorganizzati. Chi saprebbe dire qual è il prezzo di un certo volume di dati? Dipende da tante cose: dalla varietà, dal numero di azioni e di individui che lo ha generato, dallo specifico impiego attuale e prospettico[98].
Una risposta ha provato a darla il «Financial Times»[99], mettendo online, a disposizione dei suoi lettori, un vero e proprio calcolatore che consente di avere un'idea del valore del dato del singolo individuo a seconda delle caratteristiche personali. Come emerge dall'uso del calcolatore online, se da un lato, il valore del dato personale cambia al mutare delle caratteristiche dell'individuo, dall'altro, la comparazione tra persone con redditi diversi fa emergere un differenziale nel valore dei relativi dati di solo qualche decina di centesimi di dollaro. Nello stesso spirito, è stata di recente lanciata a Londra un'*app* (*ErnieApp*) frutto di una start up italiana. Attraverso una serie di indicatori l'*app* rivela agli utenti quanto valore le piattaforme digitali stanno estraendo dall'uso dei loro dati, permettendo così agli utenti di «negoziare» potenzialmente, in modo semplice e istantaneo, il livello di permessi circa la *privacy* digitale, in funzione della condivisione o restituzione di parte del valore generato[100].
Tutto ciò mostra, concretamente, l'esistenza di un potenziale mercato che caratterizza l'ecosistema dei *big data* e ne spiega il funzionamento. Ma, come ha ben chiarito il premio Nobel Ronald Coase, perché un mercato funzioni occorre (almeno) che i beni in esso scambiati abbiano diritti di proprietà ben definiti, attribuendo ai proprietari il diritto di controllo sui beni.

[97] REZZANI A., *Big data. Architettura, tecnologie e metodi per l'utilizzo di grandi basi di dati*, op. cit., p. 132.
[98] *Ibidem*, p. 133.
[99] HARFORD T., *Big data are we making a big mistake?*, in *Financial Times*, 28 marzo 2014, p. 14.
[100] *Ibidem*, p. 17.

8. Tecnologia

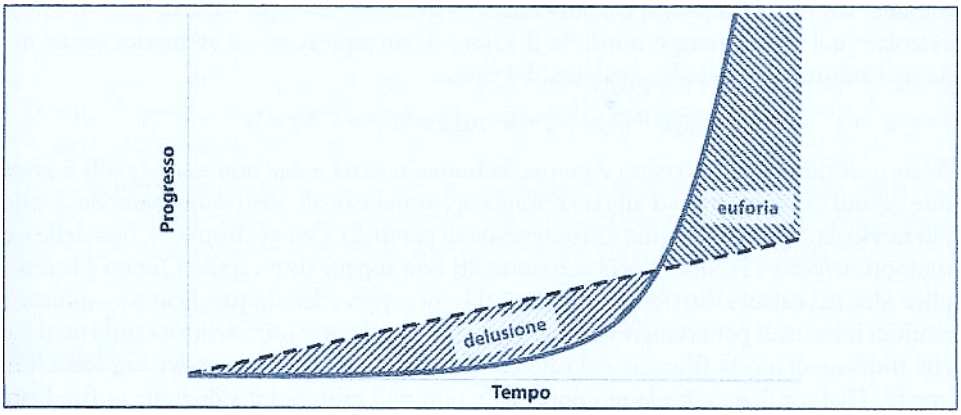
Il fattore scatenante che più di tutti può spiegare il successo esplosivo dei Big Data è di natura sostanzialmente tecnologica: è la crescita esponenziale della capacità di calcolo, trasferimento e immagazzinamento dei dati, frutto di progressi combinati in varie discipline, soprattutto nell'elettronica digitale, le telecomunicazioni e l'informatica. Esiste una regola empirica che descrive bene questa crescita tecnologica: proposta nel 1965 da Gordon Moore, cofondatore di Intel, la "legge di Moore" si è rivelata valida fino ai nostri giorni. La sua formulazione suona molto arida e tecnica: "la densità di transistor montati su chip tende a raddoppiare ogni 18 mesi". In realtà le conseguenze di questa legge empirica hanno toccato l'esperienza personale di molti di noi[101]:

- la velocità di calcolo dei microprocessori dei nostri computer continua ad aumentare costantemente (i 6 MHz del mitico processore Intel 286, fatto di 134.000 transistor, sono nulla in confronto ai 4000 MHz degli attuali processori Core i7, che contano miliardi di transistor);
- il volume delle schede di memoria per le macchine fotografiche digitali o delle chiavette USB o la memoria nei nostri smartphone cresce continuamente a parità di costo (passando dal contare le decine di MB negli anni 2000 ai milioni di MB attuali, ovvero migliaia di GB);
- la velocità di Internet nelle nostre case è passata dai 56 kb/s della connessione dial-up analogica degli anni Novanta (qualcuno ricorderà il modem analogico dall'inconfondibile suono robotico di connessione, ora di indiscutibile fascino vintage) al milione di kb/s della fibra ottica dei giorni nostri;
- analogamente, la velocità dello scambio dati via cellulare si è spostata dai 64 kb/s del 2G (per cui si usavano le sigle GSM/GPRS) ai 100.000 kb/s del 4G LTE moderno.
- La velocità con cui queste tecnologie si sono evolute ha seguito un ritmo tutto suo, impressionante. Un ritmo scandito da balzi di natura esponenziale e non lineare come siamo abituati a vedere per tutte le altre grandezze, come l'inflazione, il prodotto interno lordo o il nostro stipendio, le quali aumentano

[101] DE MAURO A., *Big Data Analytics. Analizzare e interpretare dati con il machine learning*, op. cit., p. 143.

o diminuiscono di pochi punti percentuali ogni anno. Queste grandezze tecnologiche sono invece state arricchite di diversi zeri nel giro di pochi anni[102].

Figura 1. - Progressione esponenziale (linea continua in rosso) a confronto con quella lineare (linea tratteggiata in nero). La prima all'improvviso supera di gran lunga quella lineare.



Guardando in Figura 1, osserviamo un confronto tra il tipico andamento delle progressioni lineari e quello delle progressioni esponenziali: le prime (linea nera tratteggiata) mantengono un vantaggio importante per molto tempo, fase durante la quale si rimane delusi dalla crescita delle grandezze esponenziali. A un certo punto, tutto cambia molto velocemente: la progressione esponenziale supera quella lineare e nel giro di poco tempo la rende irrilevante, attirando un grande senso di euforia. I Big Data, supportati dal progresso tecnologico spiegato dalla legge di Moore, hanno superato questo punto critico qualche anno fa e vivono ora una fase di euforia pervasiva. I dati ci sono sempre stati, così come le tecnologie per manipolarli. Anche le tecniche di intelligenza artificiale - che scopriremo insieme nei prossimi capitoli - esistevano da decenni, ma erano confinate nei laboratori e negli articoli accademici.

[102] La densità di transistor su un chip non può crescere all'infinito: quando le dimensioni dei transistor diventano troppo piccole si generano fenomeni parassiti di natura quantistica (normalmente incontrollabili) che non permettono il funzionamento corretto del dispositivo. Detto questo, la crescita esponenziale di prestazioni digitali predetta dalla legge di Moore non sembra essere a rischio. Vi sono diverse strade alternative su cui continuare a ricercare, tutte molto accattivanti, per esempio: computer quantistici (che sfruttano i fenomeni tipici della meccanica quantistica per ampliare enormemente le capacità di calcolo), circuiti integrati tridimensionali (che superano l'attuale architettura "piatta" dei microprocessori), chip neuromorflci (ispirati dal funzionamento del cervello umano), architetture parallele (come già avviene per i processori *multi-core*) e cosi via.

Poi tutto è cambiato: il punto critico è stato superato e i Big Data hanno raggiunto le nostre vite e le nostre aziende, ripetendo la progressione tipica delle tecnologie cosiddette esponenziali.

Nel presentare gli elementi tecnologici ci atteniamo a una suddivisione dei vari elementi basata sul ciclo di vita dei big data, che si compone delle fasi seguenti:

- Acquisizione (o *data ingestion*).
- Immagazzinamento e organizzazione.
- Trasformazione e analisi.

8.1 Acquisizione

L'acquisizione dei dati può essere eseguita con modalità differenti a seconda della fonte e del formato. Per i dati presenti su RDBMS il trasferimento può avvenire tramite strumenti quali Sqoop, che fa parte della piattaforma Hadoop[103]. Come alternativa vi sono strumenti quali gli ETL[104] tradizionali che si sono dotati di connettori per le tecnologie big data (Hadoop HDFS, HBase e altri database NoSQL). Per i dati prodotti in continuo e con velocità elevata, esistono gli strumenti di data streaming in grado di far fronte a tali scenari: ci riferiamo a componenti di Hadoop quali Fiume, Storm e Kafka.

La connessione con le fonti dei dati avviene attraverso driver ODBC, oppure tramite API di altro tipo messe a disposizione dal provider dei dati. Le API, *Application Programming Interface*, sono protocolli utilizzati come interfaccia di comunicazione tra componenti software. Esse consistono in insiemi di *routine*, strutture dati o variabili che permettono al programmatore di richiamare le funzionalità di un'applicazione di terze parti. Due importanti esempi sono le *Twitter API* e le *Graph API* di Facebook. Esse consentono di interfacciarsi con la piattaforma di micro-blogging e con il social network, esaminando nel primo caso tutti i *tweet* legati agli argomenti d'interesse e, nel secondo, tutti i contenuti pubblici (o accessibili tramite "amicizia") che rispondono ai criteri di ricerca desiderati. Anche i motori di ricerca mettono a disposizione API. Per esempio Yahoo ha un linguaggio *SQL-like*, per eseguire interrogazioni al proprio motore.

---

[103] WHITE T., *Hadoop: The Definitive Guide*, op. cit., p. 66.

[104] ETL è l'acronimo di *Extract Transform and Load*. Gli ETL sono strumenti utilizzati tipicamente per l'estrazione dei dati dalle fonti originali, la loro trasformazione e l'inserimento nel *data warehouse*.

8.2 Immagazzinamento e organizzazione

La conservazione dei dati pone due problemi di difficile risoluzione con le tecnologie tradizionali, quando si parla di big data: la gestione di una grandissima mole di dati, la presenza di dati non strutturati o semi-strutturati. La soluzione a tali problematiche è fornita da due tipi di tecnologie: la piattaforma Hadoop e i database NoSQL[105].

Hadoop è un software open source, affidabile e scalabile, per il calcolo distribuito, cioè il calcolo che avviene su un sistema di computer autonomi, ma collegati tra di loro in rete. I software di calcolo distribuito sfruttano la rete di computer, suddividendo su di essi l'esecuzione di operazioni. In questo modo la capacità di calcolo di ciascun elaboratore si somma a quella degli altri, consentendo di affrontare problemi complessi o su grandi quantità di dati, non risolvibili attraverso un singolo elaboratore. Hadoop nasce per essere utilizzato su ciò che è definito *commodity hardware*, cioè sistemi a basso costo che non hanno tutte le caratteristiche di affidabilità delle macchine più costose. Hadoop è infatti in grado di garantire la disponibilità del sistema e di prevenire la perdita di dati a fronte di problemi hardware. L'assenza di costi di licenza e l'abbattimento dei costi per la costruzione del cluster, costituiscono uno dei punti di forza della piattaforma.

Descriviamo in breve le componenti basilari di Hadoop[106]:

- HDFS: il file System distribuito che fornisce un'elevata capacità di accesso ai dati.
- YARN: sistema di *scbeduling* e gestione delle risorse del cluster.
- MapReduce, Tez: sistemi di *parallel processing* di grandi quantità di dati.

In Hadoop si innestano numerosi altri progetti di Apache, alcuni dei quali sono completamente integrati nel *framework* e possono funzionare solo all'interno di esso. Tali software sono talmente numerosi e interconnessi che si parla di "ecosistema Hadoop" per indicare la complessità e le interazioni presenti nella piattaforma.

Altre componenti possono lavorare in unione con Hadoop, ma non ne dipendono completamente. Un esempio è Spark, un motore di calcolo distribuito molto efficiente, che può essere installato sul cluster Hadoop, ma che può anche lavorare su altri sistemi.

Come si è detto, Hadoop è un sistema di calcolo distribuito. Affinché le operazioni di

[105] WHITE T., *Hadoop: The Definitive Guide*, op. cit., p. 68.
[106] *Ibidem*, p. 70.

calcolo avvengano è necessario che ciascun computer possa accedere ai dati. L'accesso è fornito da HDFS (*Hadoop Distributed File System*)[107], il quale, tra le altre cose, garantisce che i dati siano ridondati nel cluster[108], rendendo le operazioni sui dati stessi immuni dall'eventuale guasto di un nodo. HDFS accetta dati in qualsiasi formato, strutturato o non strutturato. Le operazioni di calcolo avvengono utilizzando la componete MapReduce, che lavora secondo il principio *divide et impera*[109]: un problema complesso, che utilizza una gran mole di dati, è suddiviso, assieme ai dati stessi, in piccole parti processate in modo autonomo. Una volta che ciascuna parte del problema è stata calcolata, i vari risultati parziali sono "ridotti" ad un unico risultato finale. L'architettura di Hadoop è molto flessibile e consente la realizzazione di motori di calcolo distribuito facilmente innestabili nella piattaforma. E il caso, per esempio di Tez, che costituisce un'alternativa a MapReduce.

HDFS è però soltanto un file System, ovvero un sistema di organizzazione dei dati e non è pensato per replicare le funzionalità tipiche di un database, come per esempio il recupero di un singolo record in tempi rapidi. Per questo nell'ecosistema Hadoop esistono applicazioni, quali HBase, che sfruttano le componenti HDFS e MapReduce, ma che sono database a tutti gli effetti.

Oltre ad Hadoop, per lavorare su dati con strutture variabili, non adatti ad una rappresentazione tabellare, vi sono numerosi sistemi che fanno parte dei cosiddetti *database NoSQL* (Cassandra, Berkeley DB, MongoDB, Neo4J). Si tratta di motori database che non aderiscono al modello relazionale. Quest'ultimo vede, invece, gli RBDMS (Relational Database Management Systems) strutturati attorno al concetto matematico di relazione, altrimenti detta tabella. Il sito *nosql-database.org*[110] offre una definizione dei database NoSQL: essi sono basi dati non relazionali, distribuite, open source e scalabili[111].

---

[107] Con il termine *File System* si intende la tecnica con sui i file sono organizzati su un supporto quale l'hard disk o il DVD.

[108] Insieme di computer (detti anche *nodi*) che formano il sistema Hadoop.

[109] Dividi e domina. E una locuzione latina con cui si indica che la divisione e la discordia dei popoli gioca a favore di chi vuol conquistarli e dominarli. Alcuni fanno risalire il motto ai tempi degli antichi greci, altri lo attribuiscono a Filippo il Macedone, altri ancora a vari imperatori romani o a Giulio Cesare. In informatica il termine è utilizzato per indicare tecniche di soluzione di problemi complessi, che vengono suddivisi in problemi più piccoli di semplice soluzione. Le soluzioni parziali sono poi combinate per ottenere la soluzione del problema iniziale.

[110] II sito offre una lunga lista di database NoSQL.

[111] La scalabilità di un sistema informatico è la sua capacità di essere facilmente modificabile nel caso di variazioni notevoli della mole di dati o di elaborazioni.

8.3 Trasformazione e analisi

Anche se concettualmente le fasi di trasformazione dei dati e di analisi si svolgono in tempi diversi e hanno scopi differenti, dal punto di vista degli strumenti utilizzati vi sono delle sovrapposizioni.

MapReduce è lo strumento nativo di Hadoop per la realizzazione di trasformazioni dei dati, calcoli e analisi. Tuttavia MapReduce è complesso e non è alla portata di tutti gli utilizzatori della piattaforma: occorre infatti essere conoscitori esperti del linguaggio Java per realizzare programmi MapReduce.

Fortunatamente esistono strumenti di più alto livello per interagire con i dati in Hadoop. Uno di essi è *Pig*, caratterizzato da un approccio procedurale e da un linguaggio, Pig Latin, che in alcuni aspetti somiglia all'SQL. Pig Latin consente di scrivere sequenze di operazioni di trasformazione in maniera piuttosto semplice: sarà infatti la piattaforma Pig a convertire i comandi nelle fasi MapReduce opportune. Pig può essere utilizzato sia per realizzare operazioni di data ingestion, sia per elaborare e analizzare i dati.

Un altro importante strumento per la preparazione dei dati è *Hive*, che è definito come il sistema di *data warehousing* di Hadoop. Hive consente di aggregare dati, eseguire *query* e analizzare grandi *dataset* utilizzando il linguaggio HiveQL, che ormai è sovrapponibile all'SQL, salvo piccoli particolari. HiveQL maschera la complessità della scrittura di funzioni MapReduce, consentendo, anche a chi non fosse esperto, di sfruttarne i meccanismi.

Per compiere analisi complesse, come per esempio l'esplorazione dei dati con tecniche di predictive analytics, occorre utilizzare strumenti diversi da Hive o Pig. Sempre da Apache proviene Mahout, una piattaforma di *machine learning* dedicata in particolare alla costruzione di *recommendation engine,* al *clustering* e alla *classificazione*.

Oltre agli strumenti completamente integrati in Hadoop, citiamo Spark, un *tool* di calcolo distribuito che racchiude in sé numerose funzionalità:

- Data ingestion (per esempio via streaming).
- Elaborazione e trasformazione dei dati.
- Analisi attraverso un'interfaccia SQL.
- Analisi avanzata tramite la libreria di machine learning.

L'utilizzo di Spark, in combinazione con Hadoop si sta diffondendo e si sta

dimostrando un'accoppiata molto efficace per la gestione dei *big data*.

Infine facciamo un cenno a strumenti esterni, che lavorano in congiunzione con Hadoop. L'integrazione può avvenire sia tramite connettori in grado di leggere e scrivere su HDFS, Hive o HBase, sia tramite un'integrazione più marcata che è in grado di sfruttare le caratteristiche distribuite di Hadoop. Tra i software che possono lavorare in entrambe le modalità citiamo R, uno dei *tool* statistici e di machine learning (ma non solo...) più utilizzati dai *data scientist*. R è un progetto *open source*.

8.3 *Industry 4.0*

L'Industry 4.0 è un tema "caldo" e si presume che lo sarà ancora per lungo tempo[112]. Non si tratta soltanto di una nuova parola alla moda, anche se, come di consueto, la si ritrova molto spesso sui materiali commerciali di software *vendor* e aziende di consulenza, più che in fabbrica. Ciò accade quasi sempre con i fenomeni che, come i big data, sono dirompenti e racchiudono un potenziale altissimo, ma richiedono tempo, investimenti e soprattutto competenze per una reale implementazione.

Com'è accaduto per i *big data*, non vi è una vera e propria definizione per l'*Industry 4.0*, anche se la possiamo descrivere come un *processo che ha il suo punto di arrivo nella produzione industriale (quasi) completamente automatizzata e interconnessa*[113]. Ovviamente tutto ciò ha diversi impatti non ancora quantificabili: quali investimenti occorre realizzare? Quali saranno i reali benefici in termini di produttività? E, infine, quali saranno le conseguenze sul mondo del lavoro?

Grandi aziende di consulenza come *Boston Consulting Group* e *McKinsey* hanno individuato alcuni fattori abilitanti dell'*Industry 4.0*. Oltre alla disponibilità di nuove generazioni di macchinari (dalla stampa 3D alle innovazioni nella robotica, delle comunicazioni *machine-to-machine*, all'*additive manufacturing*), alcuni dei principali fattori sono identificabili nelle tecnologie di elaborazione dei dati: le tecnologie *big data*, le tecnologie IOT, il *cloud computing* e gli *advanced analytics* (in particolare i sistemi basati su tecniche di *machine learning*)[114].

Vediamo ora due tipologie di attività che rientrano nell'ambito dell'*Industry 4.0* e che

---

[112] FORD M., *Industry 4.0: Making the first move*, SMT magazine, 2016, pp. 31 ss.
[113] *Ibidem*, p. 33.
[114] GOSNEY M., *5 ways industry 4.0 could change your factory*, 2015, in *www.worksmanagement.co.uk.*

sono già implementate in alcune realtà industriali. La prima è data dall'analisi, in tempo quasi reale, della situazione degli impianti, mentre la seconda è legata al concetto di *proactive maintenance* ed è comunque in stretta relazione con la prima. L'analisi in quasi *real-time*, o comunque realizzata con tempi di latenza brevi rispetto al momento in cui il dato si produce, crea alcune criticità quando il flusso di dati è particolarmente veloce e la loro quantità per unità di tempo è elevata[115].

Gli strumenti di *stream analytics* affrontano queste criticità consentendo l'acquisizione e la contestuale elaborazione di dati provenienti da fonti ad alta velocità. Diviene quindi possibile analizzare, anche con strumenti visuali, il flusso di dati provenienti, per esempio, dai sensori montati sugli impianti in modo da verificarne il corretto funzionamento. Sempre con le stesse tecniche si è in grado di avere una visione in tempo reale dei parametri di produzione. Se alla *stream analytics* si aggiungono tecniche predittive, si possono realizzare sistemi di *proactive maintenance*. Il loro funzionamento è semplice, per lo meno dal punto di vista logico. Dapprima occorre creare un modello predittivo in grado di identificare eventuali anomalie di funzionamento. In seguito, durante l'attività degli impianti, avviene l'acquisizione dei dati dai sensori. I *tool* di *stream analytics* li trasformano opportunamente, in modo che possano essere utilizzati come input dal modello predittivo. Quest'ultimo valuta se la situazione rientra nella normalità oppure se ci si trova vicini a un probabile guasto. L'identificazione preventiva dei guasti agli impianti consente di mettere in atto un fermo programmato degli impianti che, come tale, può essere gestito al meglio, evitando disagi e costi elevati dovuti agli imprevisti.

8.4 IOT - *Internet of Things*

Il concetto di *Internet Of Things* (IOT) riguarda gli oggetti dotati di sensori e di connettività che si trasformano in produttori di dati[116]. Tali *smart objects* raccolgono e inviano dati relativi alle performance del prodotto o alle modalità di utilizzo da parte del suo possessore.

Ampliando il concetto di IOT potremmo includervi quanto detto a proposito della *Industry 4.0*, dato che anche in quel caso si parla di oggetti (le macchine o le singole

[115] *Ibidem*.
[116] O'LEARY D., *Big Data, Internet of Things and the Internet of Signs*, op. cit., p. 54.

componenti) che inviano dati sul loro funzionamento[117]. Se consideriamo un'accezione di IOT più ristretta e legata solo ai prodotti consumer, possiamo citare esempi legati al mondo dei trasporti. Il primo riguarda l'utilizzo delle cosiddette scatole nere, che sono in grado di verificare numerosi parametri dello stile di guida del conducente. Le scatole nere sono installate a fronte di sconti sulle assicurazioni RCA, offrendo così un risparmio per l'assicurato e un maggior controllo sulle liquidazioni dei sinistri da parte delle compagnie. Il secondo esempio riguarda le automobili connesse, che sono in grado di inviare dati di vario genere provenienti da numerosi sensori. I casi d'uso sono molteplici e vanno dalla manutenzione proattiva alla proposta di contenuti digitali per i sistemi di entertainment, all'ottimizzazione dei percorsi, ecc.

Altri casi riguardano i sistemi di domotica e altri ancora sono caratterizzati dall'uso di sensori in agricoltura per il monitoraggio di parametri atmosferici (temperatura, umidità, ventilazione, precipitazioni,...). Un famoso parco di divertimenti utilizza braccialetti con RFID per monitorare gli spostamenti degli utenti, al fine di posizionare correttamente il personale in corrispondenza delle attrazioni più visitate in un certo momento (o previste tali tramite algoritmi di *predictive analytics*) o per rifornire adeguatamente negozi e punti di ristoro, sempre in base ai flussi previsti.

8.5 Le *smart city*

Il concetto di *smart city* prevede l'utilizzo sempre più intensivo delle tecnologie informatiche e di comunicazione per raggiungere elevati livelli di efficienza in vari settori: il consumo energetico, l'inquinamento, la mobilità e la sicurezza nelle città. Molti dei progetti hanno in comune l'utilizzo di dati provenienti da sensori o dai social network e richiedono l'uso di tecnologie in grado di contenere e manipolare un'elevata mole di dati, che possono essere sia strutturati sia non strutturati[118].

È emblematico il caso della città di Yinchuan, in Cina, che, lontana dalle ben più grandi e caotiche Pechino e Shanghai, è stata scelta per sperimentare, con gran successo, alcune tecnologie che l'hanno trasformata in *smart city*. Alcune delle soluzioni adottate a Yinchuan riguardano la gestione dei rifiuti attraverso contenitori "intelligenti" che

[117] *Ibidem*, p.56.
[118] FLEISH E., *What is the Internet of things? An economic perspective*, op. cit., p. 87.

segnalano quando sono pieni, consentendo così di programmare la raccolta nel modo più efficiente. Un'altra implementazione interessante riguarda il traffico: la città conta circa 500.000 veicoli dei quali ben oltre la metà sono stati dotati di un'etichetta RFID[119], per tracciarne i movimenti. Il sistema di gestione del traffico utilizza i segnali RFID unitamente ai dati provenienti dalle telecamere per monitorare in tempo reale e prevedere i flussi di traffico. Sulla base delle rilevazioni e delle previsioni è possibile, per esempio, calibrare i tempi dei semafori.

9. Gestione del dato

9.1 «*Big data analytics*»

Nel 1890 due giuristi statunitensi, Samuel D. Warren e Louis D. Brandeis, pubblicarono *The Right to Privacy* sulla «Harvard Law Review»[120]. Da allora, una ricchissima letteratura interdisciplinare ha riguardato i dati, intesi come *informazioni personali*. Di solito, tali informazioni digitali sono usate per finalità pubblicitarie, garantendo un formidabile *asset* ai fini della personalizzazione (ossia di una *targetizzazione* sempre più spinta) dei messaggi commerciali. Tutto ciò ha rivoluzionato l'intero ecosistema pubblicitario mondiale: perdono inevitabilmente potere economico gli editori tradizionali (coloro, cioè, che esplicitamente organizzano e veicolano un contenuto), a favore di nuovi protagonisti della diffusione di contenuti che acquisiscono la maggior quantità possibile di dati digitali (*social network*, motori di ricerca, siti di *e-commerce*, ecc.)[121].

Il valore aggiunto dell'informazione si sposta così *dal contenuto al dato*, dall'editore al nuovo intermediario (ossia la piattaforma online). Le caratteristiche dell'individuo cui sono collegati i dati raccolti riguardano aspetti socio-demografici (età, genere, etnia, livello di studio, stato matrimoniale), composizione familiare, condizioni di salute, proprietà possedute (immobiliari e mobiliari), hobby e interessi, profili di

[119] RFID (*Radio Frequency Identification*) è una tecnologia per l'identificazione automatica di oggetti, animali o persone basata su dispositivi elettronici (detti *tag*) capaci di rispondere all'interrogazione a da parte di lettori fissi o portatili a radiofrequenza. Un *tag* identifica in modo univoco un oggetto. La tecnologia RFID può trovare impiego in: automatizzazione degli inventari, bigliettazione elettronica, logistica e spedizioni, passaporti, antitaccheggio, rilevazione di parametri ambientali, controllo della temperatura degli alimenti durante le spedizioni e molto altro.

[120] WARREN S.D. - BRANDEIS L.D., *The Right to Privacy*, Harvard Law Review, Vol. 4, No. 5. (Dec. 15, 1890), pp. 193-220.

[121] *Ibidem*.

reddito, nonché attitudine al consumo.

Analizzare il mondo dei *big data* con il solo paradigma della privacy o con la lente d'ingrandimento dei dati personali rischia, tuttavia, di farci ignorare buona parte dei fattori in gioco. Infatti, i dati assumono valore economico soprattutto perché contengono *informazioni di carattere generale* e non solo personale, cioè, forniscono informazioni sugli *schemi tipizzati* di un comportamento individuale e, come tali, offrono preziose indicazioni, anche predittive, per la *big data analytics* che ci rivelerà come certe *categorie di individui* si comportano, cosa cercano, come reagiscono a determinati stimoli, e così via[122].

Scompaiono così le analisi campionarie che hanno caratterizzato l'analisi statistica, economica e sociale, del Novecento. Le informazioni sono raccolte su universi interi di gruppi di cittadini e le analisi che ne emergono, grazie a potenti algoritmi, producono sistemi predittivi, sempre più raffinati, circa le caratteristiche di tipi di individui e dei loro comportamenti (economici, socio-politici, elettorali). Tramite l'uso di tecniche di *machine learning* e computazionali (c.d. modelli psicometrici), studiosi come Michael Kosinski, creatore dell'*app* "Mypersonality", hanno realizzato algoritmi che permettono di confrontare l'accuratezza dei giudizi espressi circa la personalità degli individui con le valutazioni delle «macchine» computazionali[123].

Peraltro, paradossalmente, le predizioni sono molto più esaustive delle informazioni consapevolmente rilasciate dagli utenti. In altre parole, i modelli di *big data analytics* permettono di «ricostruire» dati personali, indipendentemente dal loro originario rilascio, rendendo del tutto superata la tradizionale classificazione tra dati personali e dati non personali. Ma anche quella tra dati strutturati e non strutturati ai fini dell'efficacia della profilazione. Una circostanza che deve farci riflettere circa l'efficacia di approcci regolatori segmentati (di tutela della privacy o di *governance* pubblica e privata nella gestione del dato) se ancorati a distinzioni nominalistiche[124].

---

[122] O'LEARY D., *Big Data, Internet of Things and the Internet of Signs*, op. cit., p. 62.

[123] È la nuova frontiera dell'intelligenza artificiale. In un interessante studio, Kosinski (con i colleghi David Stillwell e Thore Graepel) ha dimostrato come bastino pochi *like* per identificare l'orientamento politico di un soggetto (con una probabilità dell'85%); il suo credo religioso (con una probabilità dell'82%); il genere (con una probabilità del 93%); l'origine etnica (con una probabilità del 95%).

[124] Oggi non c'è più bisogno, come affermavano Warren e Brandeis, di entrare dalla porta di servizio di una casa per spiare le informazioni personali di un individuo: basta conoscere la localizzazione Gps della casa, il percorso che la persona compie per arrivare in ufficio, magari associando queste informazioni ai dati che provengono dall'account di posta elettronica.

L'utilizzo di *modelli predittivi del comportamento sociale*, resi possibili dal lavoro di potenti algoritmi su una gran massa di dati raccolti, comporta una premessa e ha svariate conseguenze. In molte circostanze la *risorsa scarsa* quindi, non è rappresentata tanto dai singoli dati personali, spesso riproducibili, quanto piuttosto dallo *stock* accumulato dei *big data*, ossia da un insieme amplissimo di dati, strutturati e non strutturati, che si aggiornano continuamente, provenendo da fonti distinte, separate e complementari, a ritmi e velocità crescenti.

Per la loro gestione (*gathering*, *storage*, *analysis*) sono necessari elevati investimenti in infrastrutture hardware e software, che conducono il sistema ad una struttura caratterizzata da elevate economie di scala e di varietà o scopo e, conseguentemente, a mercati globali dei dati assai concentrati. Ne consegue che nei mercati in cui opera l'intermediazione delle piattaforme online si osservano strutture dei costi caratterizzate da elevati costi fissi (e irrecuperabili o *sunti*), costi marginali quasi nulli e costi medi decrescenti.

In definitiva, la *big data analytics* segue spesso il percorso circolare illustrato di seguito, come nel celebre quadro di Escher, *Mani che disegnano*:

1) l'utente, anche attraverso delle «cose» a lui appartenenti, genera il dato (qualsiasi tipo di dato);
2) il dato viene acquisito e raccolto (in prevalenza dalle piattaforme digitali);
3) il dato viene, poi, aggregato ad altri dati (di solito in banche dati semi-strutturate);
4) sull'insieme di questi dati, si utilizzano tecniche algoritmiche di *big data analytics* per l'individuazione di «ideal-tipi» (segmentazione degli utenti);
5) ciascun individuo viene attribuito ad un «tipo» (in termini di caratteristiche socio-economiche);
6) l'utente (e non più i suoi dati) riceve, attraverso algoritmi di raccomandazione, servizi personalizzati e varie forme di inserzionismo pubblicitario.

I dati veicolano anche altre e più raffinate informazioni che possono essere estrapolate dagli operatori grazie alle tecniche di *big data analytics* e di intelligenza artificiale. Queste informazioni sono alla base di molte applicazioni e servizi online la cui funzione e fonte di remunerazione principale non sono, almeno direttamente, connesse all'estrazione di valore pubblicitario, bensì alla realizzazione dello svolgimento

automatico di compiti. In questo ambito, i dati vengono utilizzati per estrapolare *euristiche*, ossia procedimenti, anche non rigorosi, che consentono di prevedere o rendere plausibile un risultato. Esempi ce ne sono a bizzeffe. Questi vanno dai dati sugli spostamenti (con applicazioni che arrivano alla guida automatica) fino alle informazioni linguistiche.

Tutto deriva dal fatto che il Web è nato come un sistema non solo decentrato, ma anche di tipo cooperativo. Le informazioni sulle euristiche sono preziose perché permettono a un individuo di acquisire metodi di risoluzione di problemi che attengono agli ambiti più vari, dalle questioni di tipo casalingo (bricolage, cucina, compiti di scuola, e così via), a quelle professionali (esistono in rete molte comunità, a partire ovviamente da quella informatica, che cooperano e condividono soluzioni), fino ad arrivare a questioni che riguardano dati sensibili come quelli sanitari (analisi delle sintomatologie, diagnosi, cure mediche, e così via).

Con l'affermazione del Web, la disponibilità di enormi fonti di dati ha rivoluzionato tutte le scienze e le tecniche che ricadono sotto l'ombrello della cosiddetta digitalizzazione, come ha sottolineato un recente rapporto del *Center for Strategy and Competitiveness* di Stoccolma dal titolo evocativo: *The Substitution of Labor*. Sono diverse le aree del sapere che, pur essendo state introdotte per la prima volta subito prima o subito dopo la Seconda guerra mondiale, si stanno sviluppando enormemente e a ritmi sempre più sostenuti grazie ai dati provenienti dalla rete[125]:

- l'*automazione*, ovvero il processo attraverso il quale attività precedentemente svolte dall'uomo vengono effettuate automaticamente da tecnologie;
- l'*intelligenza artificiale*, cioè l'attività volta a rendere «intelligenti» le macchine, ossia a farle funzionare appropriatamente e con «lungimiranza» nel loro ambiente;
- il *machine learning* (che include il *deep learning*), ovvero la capacità di un computer di imparare, di modificare i propri processi sulla base delle nuove informazioni acquisite;
- la *robotica*, che si riferisce alla creazione di sistemi artificiali progettati, costruiti e realizzati per eseguire compiti o servizi per le persone, definizione

[125] VALENTE P. - IANNI G. - ROCCATAGLIATA F., *Economia digitale e commercio elettronico*, op. cit., p. 45.

che include anche il calcolo cognitivo, e quindi semplici algoritmi e software, i cd. Bot;

- il *natural language processing*, sviluppi recenti in ambito linguistico che hanno portato le tre maggiori piattaforme online a livello globale (Amazon, Apple e Google) a sviluppare e offrire assistenti digitali o virtuali (rispettivamente Alexa, Siri e Google Home), grazie a software che interpretano il linguaggio naturale e dialogano con degli interlocutori umani allo scopo di fornire informazioni o compiere determinate operazioni;
- i *feedback processing* che fanno riferimento ai dati prodotti dai giudizi umani. I commenti degli utenti su beni e servizi permettono di valutare la qualità di ciò che viene offerto in rete, e quindi di orientare le decisioni di consumo online e offline. Queste informazioni sono alla base della cosiddetta *sharing economy*, e vengono utilizzate da tutte le piattaforme in rete come *asset* fondamentale per vendere al pubblico beni e servizi (diventando esse stesse una parte del servizio offerto), per realizzare nuove attività e per migliorare i processi industriali esistenti.

9.2 Algoritmi e *match making*

Il capitalismo digitale è il capitalismo del XXI secolo. Di solito, per descriverlo, si confrontano le prime società per capitalizzazione alla borsa di New York tra la fine degli anni Novanta e i nostri giorni. Nel 1998 le prime quattro società erano Microsoft, con oltre 270 miliardi di dollari, General Electric con circa 259, Exxon Mobil con circa 172 e Royal Dutch Shell con circa 164[126].

Nel 2018, al primo posto della classifica si trova Apple, con 123.000 dipendenti (che indirettamente ha creato quasi 2 milioni di posti di lavoro negli Stati Uniti) e una capitalizzazione di borsa pari a 911 miliardi di dollari. Al secondo posto c'è Alphabet a cui fanno capo Google Ine. e altre società controllate con 789 miliardi di dollari, al terzo posto si trova Microsoft con 695 miliardi di dollari, al quarto posto Amazon, con 624 miliardi di dollari[127].

[126] QUINTARELLI S., *Capitalismo immateriale. Le tecnologie digitali e il nuovo conflitto sociale*, op. cit., p. 176.
[127] *Ibidem*, p. 177.

Il mondo è cambiato in vent'anni, segnando la cifra di un mutamento profondo nell'organizzazione capitalistica con poche grandi imprese, a livello globale, che hanno acquisito un potere significativo nell'orientare le dinamiche di molti mercati e persino le forme e i modi del dibattito pubblico online[128].

I processi innovativi basati sui dati e gli algoritmi non si fermano tuttavia ai giganti mondiali ma pervadono l'intero sistema economico e le miriadi di start up che sperimentano nuovi servizi, anche nella forma della condivisione e della *sharing economy*. In questo contesto, la parola chiave per assicurare che i benefici delle innovazioni siano diffusi nel tessuto industriale è *interoperabilità:* la possibilità cioè di realizzare innovazioni complementari cui possa accedere un pubblico vasto di utenti finali[129].

Accanto all'innovazione dal basso, c'è quindi spazio anche per politiche pubbliche volte a favorire un'innovazione aperta, a ridurre i conflitti sulla proprietà intellettuale (contemperando certezza e accesso), a promuovere nuove politiche occupazionali[130].

Secondo il *Center for Strategy and Competitiveness* di Stoccolma, molte delle funzioni che ora vengono svolte dagli uomini saranno presto automatizzate. E non parliamo soltanto di compiti routinari e manuali, ma anche di attività cognitive di tipo non routinario.

Un ruolo speciale, nel capitalismo digitale, lo hanno le grandi piattaforme digitali che hanno sviluppato una rete globale di infrastrutture per la *data economy*. Anche qui occorre tuttavia aprire il quadro: le piattaforme digitali sono moltissime, svolgono

---

[128] Emmanuel Macron, nella sua lettera ai cittadini europei del marzo 2019, si è chiesto: «Chi può pretendere di essere sovrano, da solo, di fronte ai giganti del digitale?». Siamo entrati pienamente in quello che Jonathan Haskel e Stian Westlake chiamano il «capitalismo senza capitale», un mondo nel quale i beni intangibili come l'informazione e la proprietà intellettuale sono i fattori che più spingono all'innovazione, generando servizi di largo consumo a livello globale.

[129] Per alcuni studiosi, come Mariana Mazzucato dell'University College di Londra, questo enorme impatto innovativo non è solo il risultato di menti brillanti nei garage della Silicon Valley. È anche il prodotto cumulativo di una spinta collettiva all'innovazione generata da investimenti pubblici: l'iPhone, per esempio, dipende dalla tecnologia dello smartphone che è stata finanziata con fondi pubblici, mentre sia Internet che Siri sono stati finanziati dalla Darpa del Dipartimento alla difesa statunitense; il Gps dalla Marina americana; lo schermo touchscreen dalla Cia.

[130] Il capitalismo senza capitale si caratterizza oggi per una forte redistribuzione della produzione e del reddito e, come scrivono Haskel e Westlake, per un incremento della diseguaglianza tra paesi e all'interno di essi. Come rilevato dall'Ocse in uno studio del 2016 (*Automation and Independent Work in a Digital Economy. Policy Brief on the Future of Work*) il nuovo capitalismo induce una significativa disoccupazione di breve periodo dovuta alla sostituzione di vecchie mansioni e al ritardo nella transizione a nuove professionalità. Il processo di acquisizione ed elaborazione algoritmica dei dati si sta, infatti, allargando anche al mondo delle professioni.

compiti diversi e hanno diversi criteri di interoperabilità, standardizzazione, apertura, investimenti. Per quanto sia facile accomunarle, le piattaforme globali hanno non solo modelli di business diversi ma anche un modo differente di sfruttare il dato.

Le piattaforme digitali spesso sono denominate *over the top* per il fatto che sviluppano servizi che si trovano gerarchicamente al di sopra delle infrastrutture fisiche di telecomunicazione fisse e mobili grazie alle quali accediamo alla rete[131].

Nell'ultimo anno, gli investimenti delle cinque principali piattaforme online (appartenenti a Google, Amazon, Facebook, Apple e Microsoft) hanno superato i 50 miliardi di euro. Amazon, ad esempio, ha investito circa 130 miliardi di euro, dal 2011 al 2017, nella realizzazione di reti di *data center* e nelle relative tecnologie infrastrutturali. In questo senso, il capitalismo senza capitale non è poi così senza capitali, così come il capitalismo degli *over the top* non è senza infrastrutture di rete, basandosi su infrastrutture che finora non abbiamo considerato pienamente tali.

Nell'economia dei dati, accanto alla disintermediazione delle strutture d'impresa e dei mercati tradizionali si afferma una nuova intermediazione tra diversi versanti del mercato, ad opera delle piattaforme digitali globali, fondata su una struttura tecnologica ad altissima intensità d'investimento e basata, quanto all'uso dei dati, su un'organizzazione industriale prevalentemente connotata da forme intensive di integrazione verticale. In altre parole, una volta acquisiti, i dati tendono a essere gestiti internamente, e in via esclusiva, dalle piattaforme digitali, secondo lo schema della «delega».

Nel recente libro *Matchmakers*[132], due economisti, David Evans e Richard Schmalensee, mostrano con chiarezza la novità intrinseca del mestiere delle piattaforme digitali: sono *matchmakers*, permettono, cioè, a gruppi diversi di individui di incontrarsi e traggono profitto dal valore economico generato da questo incontro, sia esso associato a uno scambio di un servizio (per esempio tra creditori e debitori) o

[131] Un recente rapporto Ocse (*Data-Driven Innovation for Growth and Well-Being*) mostra come l'acquisizione, l'analisi e la gestione dei *big data* necessitino di notevoli investimenti infrastrutturali. La mancanza di infrastrutture di reti tradizionali che caratterizza gli *over the top* non comporta, infatti, anche assenza di investimenti. Anzi. I «giganti dei dati» presentano un elevatissimo livello di capitale investito in immobilizzazioni tecnologiche e quindi una struttura dei costi caratterizzata da elevati oneri fissi e irrecuperabili (*sunks*) e bassi costi marginali assieme ad una scala mondiale di copertura della rete (e quindi dei relativi servizi).

[132] EVANS D.S., SCHMALENSEE R., *Matchmakers: The New Economics of Multisided Platforms*, Harvard Business School Pr, 24 maggio 2016, pp. 54 ss.

alla valorizzazione dell'attenzione (e quindi all'inserzionismo pubblicitario e al *programmane advertising*). I due economisti insistono sulle peculiari novità di questi modelli di business e avvertono circa gli errori o i disastri che si possono generare nel trattare i *matchmakers* come imprese tradizionali, nel campo antitrust o della regolazione. Ad esempio, se il vantaggio economico delle piattaforme deriva dal far incontrare gruppi diversi di consumatori o produttori, si comprende bene come si possa addirittura pagare uno di questi gruppi perché partecipi. La circostanza che si offrano ad utenti servizi gratuiti non sarebbe quindi un modo di attrarre l'utente per acquisirne «ingannevolmente» il dato, ma un necessario incentivo economico per realizzare una transazione economica dii mutuo vantaggio per tutti i partecipanti ed un preciso modello) di business.

La struttura di mercati multi-versante, d'altra parte, non è peculiare soltanto dei nuovi mercati digitali basati sull'estrazione del dato (*data driven*). Essa interessa, in diverso grado, molti mercati, tra i quali quelli dei media tradizionali (Tv, quotidiani, periodici, radio).

Ad esempio, l'editore di un giornale che venda il proprio quotidiano ai consumatori e al contempo i «contatti» pubblicitari agli inserzionisti. Così, quando un editore di un quotidiano decide il prezzo di vendita al pubblico, lo fa anche considerando gli effetti che ne derivano sul collegato versante pubblicitario, ossia sul prezzo con cui vende le inserzioni commerciali sulla stessa testata.

Un esempio «storico», chiaramente esplicativo della necessità di valutare il mercato, e le relative condotte aziendali, complessivamente in tutti i versanti, riguarda la «guerra di prezzo dei giornali britannici»[133].

[133] All'inizio degli anni Novanta in Gran Bretagna il prezzo dei quotidiani «di qualità» era abbastanza stabile e uniforme: il «Times», l'«Independent» e il «Guardian» costavano 45p e il «Daily Telegraph» 48p. Nel settembre del 1993, il «Times», del gruppo Murdoch, tagliò improvvisamente il prezzo di un terzo del proprio valore (da 45p a 30p).
Ciò innescò una guerra commerciale, che si riflesse, anche se con un certo ritardo, sui prezzi di «Independent», «Daily Telegraph» e «Guardian». L'«Independent» sostenne di essere vittima di un tentativo di estromissione dal mercato in quanto l'abbassamento del prezzo del «Times» sarebbe equivalso a una perdita dei profitti di breve-medio periodo, e aveva pertanto una valenza esclusivamente anticoncorrenziale. Tuttavia, avvenne esattamente il contrario: stante la riduzione del prezzo di copertina, le vendite del «Times» aumentarono sensibilmente, così che il gruppo Murdoch potè incrementare il prezzo della pubblicità sulla propria testata, con un effetto netto positivo sui profitti complessivi. In definitiva la crescita del mercato pubblicitario sui quotidiani richiedeva un mix di finanziamento diverso. D'altra parte, questi mercati vengono denominati «a due o più versanti», proprio per la presenza di un intermediario che ponendosi al centro degli scambi mette in contatto soggetti diversi.

Le piattaforme online agiscono allo stesso modo, definendo il giusto mix tra i prezzi nei vari versanti di mercato. Grazie alla profilazione e alla granularità del dato, nonché alla capacità predittiva degli algoritmi, oggi l'«attenzione» degli utenti ha molto più valore nel caso del *programmatic advertising*. In molti casi, in un versante, al fine di acquisire il maggior numero di utenti possibili, viene fissato il prezzo uguale a zero, e i ricavi provengono direttamente dai versanti collegati.

Ciò che in passato ha strutturalmente segmentato i mercati mondiali in ambiti geografici circoscritti, ossia le diversità nazionali (o addirittura locali) dei consumatori (per gusti, lingue, culture, ecc.) nonché i limiti tecnologici per produrre, offrire e distribuire prodotti e servizi, è improvvisamente venuto meno.

Esclusi rari casi, tutto ciò è dovuto, in massima parte, all'operare di *effetti* o di *esternalità di rete* che rendono «naturalmente» internazionale anche la dimensione dei mercati dei servizi Internet.

In generale, si parla di effetti di rete (diretti) quando il valore di un bene o di un servizio per un individuo aumenta direttamente all'aumentare delle persone che posseggono il medesimo bene o aderiscono al medesimo servizio. In sostanza, l'utilità dell'utente è direttamente influenzata dalla numerosità degli altri utenti interconnessi, così che il valore, economico e finanziario, della rete cresce direttamente al crescere del numero di utenti attivi.

Gli effetti di rete assumono forme differenti tanto che si è arrivati a contarne ben 13 tipi, ma la relazione principale rimane la stessa e correla la dimensione di una rete (ad esempio un social network) al suo valore economico e sociale. L'incremento della base utenti di una rete, che comporta una crescita della disponibilità a pagare di ciascun utente, non necessariamente implica un incremento del prezzo di adesione al sistema. Infatti, anche in presenza di potere di mercato del gestore della piattaforma, quest'ultimo potrebbe rinunciare a incrementare i prezzi, allo scopo di perseguire una strategia di rapida acquisizione della base utenti, necessaria a raggiungere quella dimensione del mercato (*massa critica*) indispensabile per l'affermazione del proprio sistema.

I *social network*, ad esempio, come ha osservato Chris Geddes, devono superare una certa percentuale di penetrazione sulla propria popolazione di riferimento (il *target*) per riuscire ad affermarsi. Al tempo stesso, mentre i costi della piattaforma tendono a

crescere al più linearmente, i potenziali ricavi aumentano seguendo una legge di potenza (c.d. *power law*).

L'insieme di queste caratteristiche dei servizi a rete ha condotto Thomas Eisenmann della *Harvard Business School* a coniare l'espressione *the winner takes all* (Wta): «chi vince prende tutto». L'andamento storico della quota di Google nel mercato dei motori di ricerca (stima del mercato Usa da fonti varie - Sew, NetApplications, StatCounter) mostra in modo abbastanza eloquente sia l'esistenza di una massa critica (attorno al 30% degli utenti), sia la dinamica competitiva che caratterizza i mercati in cui il vincitore prende tutto.

D'altra parte, non è detto che si vinca tutto «per sempre». E, in un certo senso, la vicenda del vecchio gigante Yahoo!, scalzato da Google, mostra l'importanza dell'innovazione nel determinare il successo. Anche il motore di ricerca di Google, come già Yahoo!, in teoria, può essere superato da un nuovo concorrente. Ma c'è anche chi sostiene che il divario tra Google e tutti gli altri si sia ormai strutturato e sia difficilmente colmabile. Meccanismi del tipo *the winner takes all* portano a una naturale posizione di leadership sul mercato difficilmente contendibile, anche in ragione del fatto che, dal lato della domanda, al crescere degli effetti di rete, cresce anche il costo-opportunità di abbandonarla da parte degli utenti. Al riguardo è stata coniata l'espressione *tipping point*: un punto di massa critica superato il quale la piattaforma è preferita proprio per gli effetti di rete che essa genera.

Gli economisti Evans e Schmalensee hanno tuttavia criticato l'applicazione del concetto *the winner takes all* a determinati servizi online in ragione dei bassi (se non nulli) costi sostenuti dagli utenti per cambiare piattaforma (*switching costs*). In un mercato a due versanti, se un numero significativo di utenti si sposta verso una nuova piattaforma, anche gli inserzionisti pubblicitari potrebbero decidere di farlo. Inoltre, la possibilità di usare contemporaneamente più servizi dello stesso genere in piattaforme diverse (*multihoming*) - ad esempio, con l'uso congiunto di due social network, come Twitter e Facebook o di due motori di ricerca come Google e Bing - comporterebbe la possibilità di fare esperienza di servizi diversi in concorrenza tra loro (sia sotto il profilo del servizio che avendo riguardo alla concorrenza nel tempo di attenzione), un pò come si fa con lo *zapping* tra diversi canali televisivi.

9.3 Un caso interessante: II settore *banking*

Vediamo ora un caso interessante di utilizzo dei big data. Partiamo dall'uso più vicino alle tecniche tradizionali, che riguarda un esempio preso dal settore bancario. Alcune grandi banche italiane si stanno dotando (o si sono già dotate) di sistemi *big data*, *Hadoop* e *Spark* in particolare.

Molto spesso tali sistemi non sono approntati per applicazioni particolarmente sofisticate, ma per realizzare analisi che, pur basandosi su dati e tecniche tradizionali, incontrano limiti tecnici a causa dei ragguardevoli volumi di dati. Il problema è sostanzialmente quello di estendere le analisi, che già da tempo sono realizzate su dati aggregati (saldi mensili, per esempio) oppure su profondità storiche limitate (6 mesi/1 anno), a dettagli più granulari o a orizzonti temporali molto più estesi. Le grandi banche, con milioni di clienti e decine di milioni di rapporti (cioè conti correnti, mutui, titoli, ...) vedono prodursi nei loro sistemi operazionali parecchi milioni di transazioni giornaliere, la cui sintesi (mensile, per esempio) può far perdere molte informazioni preziose che si celano nelle singole operazioni.

Attività di tipo tradizionale su tali volumi potrebbero quindi essere ricondotte alla realizzazione di sistemi di *business intelligence* su periodi temporali molto più ampi, con la capacità di effettuare operazioni di *drill-down*[134] sui dati di estremo dettaglio. Più semplicemente vi può essere l'esigenza di avere sempre disponibili dati del passato sia per scopi analitici, sia per imposizioni normative. La reperibilità dei dati meno recenti, soprattutto con i sistemi *legacy*, è spesso difficoltosa e richiede tempistiche misurabili in giorni. Un'architettura basata su Hadoop, consentirebbe di mantenere un'adeguata storicizzazione delle transazioni, lasciando la possibilità di interrogarle con tempi di latenza molto ragionevoli (minuti o decine di minuti).

In altri casi la mole dei dati impedisce la realizzazione e l'esecuzione di modelli predittivi in tempi ragionevoli. L'impiego di tecnologie come Spark ha consentito, in quelle situazioni, di realizzare modelli di *churn*, *clustering* o *campaign targeting*[135] su

[134] Tra le operazioni rese possibili dagli strumenti di Business Intelligence vi è il *drill-down*. Esso consente di passare da un livello di maggior aggregazione ad un livello di minor aggregazione, visualizzando via via dati di maggior dettaglio che, naturalmente, devono essere disponibili nel database sottostante al sistema di BI.

[135] Si tratta di tecniche di *predictive analytics* volte a individuare i clienti a rischio di abbandono (*modelli di churn*), oppure i clienti con alta propensione all'acquisto (*modelli di targeting*). I *modelli di clustering* invece raggruppano la clientela in insiemi omogenei, in modo da condurre su di essi ulteriori analisi.

una base dati di notevoli dimensioni, con tempi di esecuzione dei calcoli dell'ordine di poche ore o, in certi casi, di qualche decina di minuti.

Un esempio di analisi più innovativa è legato all'utilizzo di strumenti di elaborazione del testo per recuperare informazioni anche dai campi descrittivi. In particolare ci riferiamo alle causali in formato libero che è possibile inserire quando si emettono bonifici tramite le applicazioni di internet banking. Alla base di tale tipologia di analisi vi è l'esigenza di sfruttare ogni informazione per aumentare la conoscenza della propria clientela, al fine di offrire prodotti e servizi più adatti ai singoli clienti o di prevenire dinamiche di *churn* che non sono infrequenti in ambito bancario.

CAPITOLO SECONDO

ETICA E REGOLE NELL'USO DEI DATI

1. Pluralismo e democrazia

Nel 2014, nel suo messaggio per la "Giornata Mondiale delle Comunicazioni", Papa Francesco ha scritto: «Internet può offrire maggiori possibilità di incontro e di solidarietà tra tutti, e questa è cosa buona, è un dono di Dio».

Queste sorprendenti parole si applicano all'ecosistema digitale nel suo complesso. Le inedite «possibilità di incontro», infatti, non sono rese possibili solo da Internet ma da tutte le applicazioni digitali che stimolano la partecipazione, il rilascio e lo scambio di informazioni.

Le domande, le perplessità, le criticità che riguardano l'attuale evoluzione della società digitale non possono mettere in dubbio le straordinarie conquiste raggiunte dalla rivoluzione digitale.

E la promessa che essa ci offre: una nuova dimensione pubblica e privata della nostra esistenza che trasforma, arricchendolo, lo spazio delle nostre libertà economiche, sociali, civili. La rete - ha scritto qualche anno fa Lawrence Lessig[136] - permette di esportare il primo emendamento degli Stati Uniti in tutto il mondo e, con esso, l'affermazione della libertà di espressione, del *free speech*, contro ogni tentativo di ostacolo o limitazione da parte del potere pubblico.

Nondimeno, dobbiamo comprendere i rischi che abbiamo di fronte, anche con riferimento al potere privato e al connubio tra poteri pubblici e privati, identificarli e superarli, senza tuttavia commettere l'errore - quello sì fatale - di pensare che tutto ciò che si manifesta davanti a noi sia, per ciò stesso, il migliore dei mondi possibili. Che la migliore policy digitale sia nessuna policy e che il *business as usual* sia l'unico modo di mantenere gli indubbi benefici che la società digitale, anzi Dio, ci ha donato.

Fa dunque riflettere, e come non potrebbe, un articolo-manifesto di Hossein Derakhshan, ricercatore presso lo *Shorenstein Center* di Harvard e il *Mit Media Lab*,

[136] LESSIG L., *Cultura libera. Un equilibrio fra anarchia e controllo, contro l'estremismo della proprietà intellettuale*, Apogeo, Milano, 2005, pp. 12 ss.

apparso un anno dopo il messaggio del Papa, su Medium.com, dal titolo: *The Web We Have to Save* (*Il Web che dobbiamo salvare*). Derakhshan è noto soprattutto per essere stato un attivista politico e un blogger in Iran e per avervi trascorso sei anni in prigione a causa della sua attività. In quel saggio, il *blogfather* spiega quanto sia cambiato il Web nei sei anni passati in prigione: «nel 2008, non c'erano Instagram, SnapChat, Viber, WhatsApp. Al loro posto c'era il Web, e sul Web c'erano i blog: il posto migliore al mondo nel quale trovare pensieri, notizie e analisi diverse e alternative». Nei blog, aggiunge Derakhshan, i collegamenti ipertestuali, gli *hyperlink*, garantivano diversità e decentralizzazione e rappresentavano lo spirito aperto e interconnesso del Web. Sia perché permettevano di arricchire i testi e di verificare in proprio la correttezza delle fonti, sia perché permettevano a diverse fonti di animare le proprie pagine Web in modo decentralizzato. Oggi, secondo Derakhshan, le pagine Web al di fuori dei social network stanno morendo.

Nel febbraio del 2017, su «Scientific American», fu pubblicato un altro manifesto, scritto da un gruppo di eminenti studiosi di varie discipline, tra i quali Bruno Frey, dal titolo: *Riuscirà la democrazia a sopravvivere ai big data e all'intelligenza artificiale?*. L'articolo sottolineava, tra gli altri rischi, quello politico-elettorale.

Un anno dopo, nel marzo del 2018, Sir Tim Berners-Lee, il creatore del Web, scrisse una lunga lettera aperta in occasione del ventinovesimo compleanno del World Wide Web. Anche qui il titolo era un grido d'allarme: *Il Web è sotto minaccia. Unisciti a noi per combatterla*.

L'appello veniva dallo stesso uomo che seppe emozionare il mondo durante la cerimonia di apertura dei giochi olimpici di Londra 2012. Quella sera d'estate, per celebrare la nascita del Web, Berners-Lee scrisse un *tweet* che fu proiettato sul pubblico dello Stratford Stadium: *this is for everyone*. Il Web per tutti e per ciascuno di noi[137].

Che cosa è successo a Berners-Lee e a tutti gli altri? Com'è stato possibile che quello

[137] Gli stessi concetti sono poi stati ripetuti nel grande evento organizzato al Cern, per la celebrazione dei trent'anni del Web, 1'11 marzo 2019. Secondo Berners-Lee le tre minacce dalle quali difendere il Web sono: *a)* gli intenti dolosi premeditati, come la pirateria e gli attacchi informatici promossi dagli Stati, comportamenti criminali e molestie online; *b)* la struttura del sistema crea incentivi perversi, in cui il valore d'uso è sacrificato, ad esempio modelli di introiti basati sulla pubblicità, che premiano a livello commerciale il *click bait* e la diffusione virale di disinformazione; *c)* le conseguenze involontarie negative derivanti da buone intenzioni, come i toni indignati e la polarizzazione del dibattito online.

che era «un dono di Dio» diventasse, nel giro di pochi anni, una minaccia così spaventosa? Che fine ha fatto la primavera araba dei social nella piazza di Tahir e cosa c'entrano i *big data* e gli algoritmi con la minaccia alla democrazia?

Nella sua lettera aperta, Berners-Lee si chiedeva come fosse stato possibile che quella che una volta era una piattaforma aperta, nella quale trovava spazio una vasta selezione di blog e siti, potesse esser stata schiacciata sotto il peso di alcuni soggetti dominanti che divorano i rivali di più piccole dimensioni e «che si possono contare sulle dita di una mano», che possono «controllare quali idee e quali opinioni sono viste e condivise» e che «sono capaci di proteggere la loro posizione creando barriere all'entrata per i concorrenti». «Negli ultimi tempi - scrive Berners-Lee - abbiamo visto teorie cospiratorie diventare virali sui social, account falsi creare tensioni sociali, attori esterni influenzare le elezioni e criminali rubare un tesoro di dati personali»[138].

Jeff Bezos, il Ceo di Amazon, durante la conferenza Wired 25 a San Francisco ha affermato che «Internet nella sua attuale incarnazione è una macchina che conferma i pregiudizi. Hai un determinato punto di vista, fai una ricerca su Web, trovi conferma del tuo punto di vista».

Insomma, da più parti emerge come l'oceano-Web nato per soddisfare la nostra libertà di ricerca, per navigare liberamente sul mare delle idee, si riveli pieno di banchine, rotte prestabilite e porti che sono, allo stesso tempo, il risultato della nostra libertà e il confine delle nostre scelte. Inutile negare che tutto questo recente pessimismo abbia anche a che fare con ragioni politiche, essendosi il dibattito alimentato a seguito delle nuove forme di *propaganda algoritmica* che, sperimentate da Barack Obama, sono poi esplose con la Brexit e l'elezione di Trump. A seguito di questi eventi, in particolare, si sono moltiplicate le accuse di influenze e manipolazioni sul voto in molti paesi da parte di gruppi e organizzazioni straniere. Ma anche con il dilagare delle parole d'odio sul Web e sui social media, la morte del «politicamente corretto», il risorgere di pulsioni di antisemitismo e razzismo.

Il 4 marzo 2019, nella sua lettera inviata ai cittadini europei, il presidente della Repubblica francese, Emmanuel Macron, ha insistito su un'insidia, che a suo avviso «minaccia tutta l'Europa: gli sfruttatori della collera, sostenuti dalle false informazioni,

[138] Berners-Lee T., *L'architettura del nuovo web: dall'inventore della rete il progetto di una comunicazione democratica, interattiva e intercreativa*, Feltrinelli, Milano, 2018, p. 43.

promettono tutto e il contrario di tutto».

Come hanno scritto Roger Eatwell e Matthew Goodwin (*National Populism*, 2018) tutto ciò è soprattutto una reazione, di intere generazioni deluse o tradite, alle paure della globalizzazione e alle nuove povertà, diseguaglianze e incertezze generate dalla profonda crisi economica finanziaria mondiale del 2008. Dobbiamo avvicinarci a questo dibattito con le giuste dosi di precauzione, cautela, disincanto, neutralità. Dobbiamo cioè evitare eli cadere noi stessi proprio in quei pregiudizi di conferma che vogliamo studiare, comprendere e descrivere.

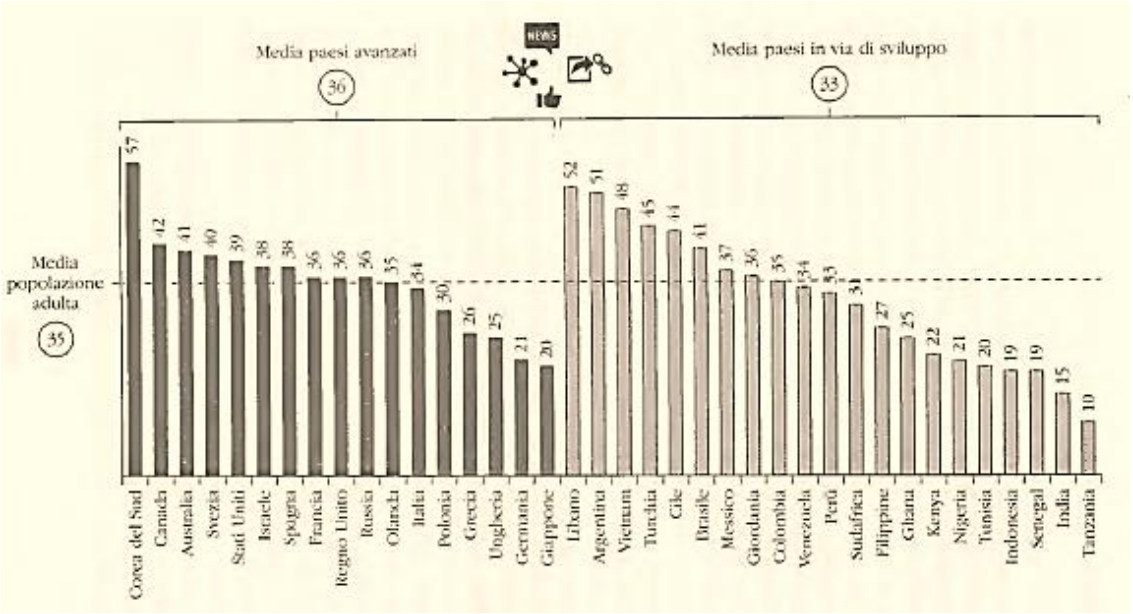
1.1 Il consumo d'informazione

Per comprendere la rilevanza del Web e delle piattaforme digitali nel design dello spazio informativo occorre capire come la profilazione algoritmica contribuisca a selezionare contenuti determinanti per la formazione dell'opinione pubblica e per l'*agenda setting* politica. La diffusione delle piattaforme online e, in particolare dei social network (non a caso ricompresi nel concetto ampio di social media), ha comportato il passaggio da un modello di integrazione verticale delle diverse fasi della catena del valore, tipica dell'editoria offline al cui interno l'editore esercitava un controllo (diretto o indiretto), ad una separazione dei diversi stadi del processo produttivo. Cambiano le informazioni e l'accesso alle informazioni, consentendo agli stessi utenti di partecipare alla produzione e riproduzione di contenuti informativi e generando una contrazione dello spazio di esercizio del ruolo di intermediario svolto dall'editore tradizionale di giornali, radio e Tv.

La nostra dieta informativa quotidiana è ormai caratterizzata da un crescente fenomeno di *crossmedialità*, cioè di uso congiunto di mezzi d'informazione tradizionali e di varie modalità offerte dal Web, soprattutto attraverso motori di ricerca come Google e Yahoo! e social network come Facebook e Twitter (fig. 3). In Italia, secondo l'ultimo rapporto di Agcom sul consumo d'informazione, questo fenomeno interessa oltre i tre quarti della popolazione. La vecchia televisione resiste (circa 1'8% della popolazione si informa solo attraverso la Tv), mentre il 5% circa degli italiani non si informa affatto. I media tradizionali (*mainstream*), come la Tv, rappresentano ancora il principale mezzo utilizzato per informarsi, sebbene l'attenzione si ripartisca in modo difforme tra le diverse generazioni.

I quotidiani sono consultati per informarsi tutti i giorni da meno del 20% di individui, mentre sempre più persone si affidano alla rete per reperirvi informazioni o semplicemente ricevono, anche senza sollecitazione, dal Web le informazioni di cui dispongono.

FIG. 3. Utilizzo dei social network per informarsi in Italia e nel mondo (dati percentuali).



Oltre un quarto della popolazione italiana reputa Internet lo strumento più importante per informarsi. La radiografia italiana non si discosta, in questo, da quella degli altri paesi avanzati. Il 55% della popolazione italiana consulta almeno una piattaforma online per informarsi, mentre la fruizione delle fonti editoriali online si ferma al 39%. In particolare, social network e motori di ricerca raggiungono le porzioni più ampie di popolazione, pari ciascuna al 37%. Anche quando la finalità informativa è politica/elettorale, il reperimento di notizie e punti di vista su Internet passa in maniera prioritaria attraverso le piattaforme online, piuttosto che attraverso i siti Web o le *app* di quotidiani o altri siti di informazione online: 28% a fronte dell'8% della popolazione maggiorenne. Circa il 20% della popolazione indica una fonte algoritmica come la più importante all'interno della propria dieta informativa: motori di ricerca e social network rappresentano, rispettivamente, la terza e la quarta fonte informativa reputata come la più importante per informarsi, considerando la totalità dei mezzi di comunicazione (tradizionali e online).

In considerazione della crescente diffusione di dispositivi tra il pubblico e del moltiplicarsi delle occasioni di fruizione, cambiano i modelli di consumo dei media e dell'informazione. Dal lato dell'offerta di informazione, la maggiore disponibilità di fonti informative aumenta il cosiddetto pluralismo esterno e amplifica la libertà di informarsi, di confrontare fonti, di formarsi un'opinione autonoma. La stessa offerta è arricchita dai contenuti auto-prodotti, dalle ricostruzioni dirette di testimoni, da notizie e informazioni che non provengono soltanto da fonti giornalistiche ma da un universo poliedrico e variegato. Come ha scritto Sunstein, ciascuno può costruirsi online il proprio «palinsesto», il *daily me* quotidiano[139]. Al tempo stesso, questa ritrovata autonomia produce una qualità complessiva che è il frutto esclusivo delle nostre attitudini[140].

Si comprende allora perché un simile contesto sia naturalmente fertile e predisposto alla diffusione di strategie di disinformazione basate su notizie false e messaggi polarizzanti. Il superamento della vecchia Tv, del palinsesto per tutti, ci ha fin qui portato maggiore libertà ma, paradossalmente, minore pluralismo e maggiore polarizzazione. Un tema tanto più preoccupante quanto minore è il ricorso a fonti informative tradizionali, quali la stampa, dotate di specifiche forme di responsabilità editoriale, da parte di lettori e di inserzionisti pubblicitari. Come ha scritto nel lontano 1998 Varian, è la legge di Gresham applicata all'informazione, per cui la moneta cattiva scaccia quella buona: «l'informazione a basso costo e di bassa qualità su Internet può creare problemi a coloro che producono informazione di qualità»[141].

E così, quella che ci sembrava una finestra sul mondo, la nostra finestra protetta, finisce per essere uno specchio distorto che ci rimanda indietro l'immagine di ciò che pensavamo del mondo, prima di informarci su di esso.

La società digitale vive un paradosso: minimizza il costo di transazione nell'acquisizione di informazioni rispetto al passato ma, al tempo stesso, riduce anche l'attività di ricerca delle informazioni rilevanti. Un fenomeno rafforzato da quella che Derakhshan ha definito la morte degli *hyperlinks*. Insomma, nel Web, chi cerca trova, ma chi trova, continua a cercare? Sembrerebbe proprio di no. Ciò per due ragioni

[139] SUNSTEIN C.R., *Conformity: The Power of Social Influences*, NYU Press, 2019, p. 15.
[140] *Ibidem*, p. 16.
[141] VARIAN H.R., *Information Rules: A Strategic Guide to the Network Economy*, Harvard Business School Press, 1998, p. 45.

principali: innanzitutto, abbiamo troppe informazioni e ci richiederebbe molto tempo il leggerle, filtrarle, selezionarle. Si chiama *information overload*, il sovraccarico informativo per il quale troppa scelta ci confonde e troppa ambiguità ci farebbe sorgere nuove domande, quando sono le risposte quelle che cerchiamo. In un recente studio, Petra Persson della *Stanford University*, mostra come la concorrenza nell'offerta d'informazione, in un contesto nel quale l'attenzione è limitata, riduce la conoscenza dei consumatori, causando *information overload*[142]. Lo aveva già detto, nel 1971, il Nobel Herbert Simon che «una ricchezza di informazioni crea una povertà di attenzione e la necessità di allocare efficientemente quell'informazione tra le sovrabbondanti fonti che possono consumarla».

E anche il filosofo israeliano Yuval Noah Harari ha sottolineato come «in un mondo inondato da informazioni irrilevanti, la chiarezza è potere»[143]. La nostra attenzione è scarsa, come il tempo che dedichiamo alla ricerca di informazioni. Finiamo così per scendere alla prima fermata (*one stop shop*) del nostro viaggio alla ricerca di informazioni, con il bagaglio informativo che abbiamo raccolto in pochi click[144]. Come tutti i bagagli fatti in gran fretta, esso contiene solo le cose che ci servono. E ci serve quello che conferma la nostra visione del mondo. Nel selezionare le informazioni rilevanti siamo, infatti, attratti da quelle che più soddisfano i nostri *a priori*, cioè le nostre convinzioni di partenza: è una distorsione cognitiva che prende il nome di «pregiudizio di conferma» (*confirmation bias*)[145].

Questa *scorciatoia mentale* è la seconda ragione che riduce la nostra ricerca di informazioni, accanto al sovraccarico informativo. A volte lo facciamo consapevolmente, più spesso è l'algoritmo dei motori di ricerca o dei social network a suggerirci ciò che è probabile possa interessarci, in base al nostro profilo rivelato da scelte passate o alle scelte dei nostri amici. E l'esperienza di gruppo (*groupthink*) rafforza le nostre credenze e le nostre distorsioni cognitive, polarizzandole verso una visione ancora più estrema. E l'estremismo, poi, si rafforza con l'ignoranza, secondo un paradosso noto come *effetto Dunning-Kruger*, perché tendiamo a sopravvalutare le nostre conoscenze e competenze. Molti esperimenti, citati da Steven Sloman e Philip

[142] PERSSON P., *Attention Manipulation and Information Overload: Barriers to Consumer Protection*, Behavioral Public Policy, 2018, p. 78.
[143] HARARI Y.N., *21 lezioni per il XXI secolo*, Bompiani, Milano, 2018, p. 65.
[144] *Ibidem*, p. 67.
[145] *Ibidem*, p. 68.

Fernbach, nel libro *The Knowledge Illusion*[146], mostrano come meno sappiamo di un argomento e più ci aggrappiamo alle nostre convinzioni, diffidando di chiunque le metta in discussione, polarizzando la nostra opinione.

Come ha ben scritto Rob Brotherton «il pregiudizio di conferma entra in gioco non appena ci soffermiamo su un'idea»[147].

Non sono attitudini nuove. Ne aveva già scritto, nel 1922, Walter Lippmann nel suo saggio magistrale su *L'opinione pubblica*: «il modo in cui vediamo le cose è una combinazione di quello che c'è e di quello che ci aspettavamo di trovare. Il cielo non è lo stesso per un astronomo e per una coppia di innamorati. [...] Se non stiamo molto attenti, tendiamo a figurarci tutto quello che ci sembra conosciuto con l'ausilio di immagini già presenti nella nostra mente»[148].

Si tratta eli atteggiamenti mentali, individuali e collettivi che sono sempre esistiti, ma che nella proiezione digitale assumono una dimensione (di frequenza, di ripetizione, di interazione) tale da poter cambiare, o almeno influenzare, il campo di gioco. Secondo Evgenij Morozov: «la buona notizia è che non stiamo precipitando in un nirvana globalizzato dove tutti mangiano da McDonald's e guardano gli stessi film [...]; la cattiva notizia è che la politica globale rischia di diventare ancora più complessa, controversa e frammentata sotto la pressione delle forze religiose, nazionaliste e culturali riattizzate da Internet»[149].

Il problema è che la buona e la cattiva notizia di cui ci parla Morozov sono generate dallo stesso fenomeno - la riduzione dei costi transattivi di ricerca delle informazioni sul Web - e coesistono in un nodo gordiano, per cui cancellare la cattiva notizia comporta distruggere anche quella buona[150]. Vale per l'acquisto di prodotti e di servizi, ma vale anche per la formazione di un'idea o la condivisione di una verità fattuale. Il che però può comportare non solo la moltiplicazione di universi paralleli chiusi, ma anche la sopravvivenza di idee e istanze antistoriche, di propaganda, fatti e opinioni non veritieri che così resistono alla falsificazione e alla «naturale» selezione della

---

[146] SLOMAN S. - FERNBACH P., *The Knowledge Illusion: Why We Never Think Alone*, Riverhead Books, 2018, p. 45.

[147] BROTHERTON R. - GIACONE G. L., *Menti sospettose. Perché siamo tutti complottisti*, Bollati Boringhieri, Torino, 2017, pp. 43 ss.

[148] LIPPMANN W., *L'opinione pubblica*, Donzelli Editore, Roma, 1992, pp. 23 ss.

[149] MOROZOV E., *Internet non salverà il mondo. Perché non dobbiamo credere a chi pensa che la Rete possa risolvere ogni problema*, Mondadori, Milano, 2014, pp. 56 ss.

[150] *Ibidem*, p. 60.

storia.

Il Web rende disponibili per tutti un'enorme massa di informazioni, offrendoci un accesso democratico e universale mai registrato nella storia dell'umanità.

Di fronte al sovraccarico informativo selezioniamo attraverso un *doppio filtro*: il pregiudizio di conferma da un lato e la profilazione algoritmica dei nostri dati dall'altro.

Due economisti, Fabio Sabatini e Francesco Sarracino, hanno rilevato empiricamente che nelle discussioni online con sconosciuti, le persone indulgono con maggior facilità in comportamenti aggressivi e irrispettosi e che l'uso di social network riduce, anziché alimentare, rapporti di fiducia[151].

Ce n'è abbastanza per porsi, almeno, delle domande. Perché la sfiducia alimenta l'isolamento, sostituisce i fatti con le emozioni (inaugurando quello che è stato definito il paradigma della postverità), rafforza i pregiudizi, induce, come ricorda Sunstein, polarizzazione ed estremismo, ci rende tutti sospettosi e complottisti, alimentando discorsi d'odio, e con essi, purtroppo, crimini d'odio.

Negli ultimi anni si sono moltiplicati, a livello mondiale, studi empirici che documentano la forza trainante dell'attenzione selettiva degli utenti del Web, esclusivamente diretta verso quei soli contenuti che confermano i propri *a priori* e cancellano, dimenticano, sottovalutano tutto ciò che li falsifica[152].

Secondo l'analisi empirica pubblicata da Walter Quattrociocchi[153], uno dei massimi esperti sul tema, i modelli di consumo informativo e l'interazione degli utenti con le notizie sulle piattaforme online sono caratterizzati da una forte tendenza alla polarizzazione, all'esposizione selettiva, all'omofilia, e all'insorgenza di camere d'eco[154]. Gli utenti analizzati tendono, cioè, a selezionare le informazioni che sono coerenti con il proprio sistema di preferenze e convinzioni, formando gruppi polarizzati di persone con idee simili su narrazioni condivise, in cui le informazioni

---

[151] Sulla stessa linea, Justin Cheng e colleghi della *Stanford University* hanno pubblicato i risultati di un esperimento, dal titolo emblematico: *Anyone could become a troll*. L'uso dei social e di un certo registro di comunicazione, fatto di reciprocità nell'aggressività, ci trasforma, di solito in peggio.

[152] Un recente rapporto dell'Autorità per le garanzie nelle comunicazioni, ha per titolo *News vs Fake nel sistema dell'informazione* (2018) e misura empiricamente da un lato la rilevanza del pregiudizio di conferma (*confirmation bias*) nei social media e dall'altro la sua capacità di generare polarizzazione delle idee e del dibattito pubblico.

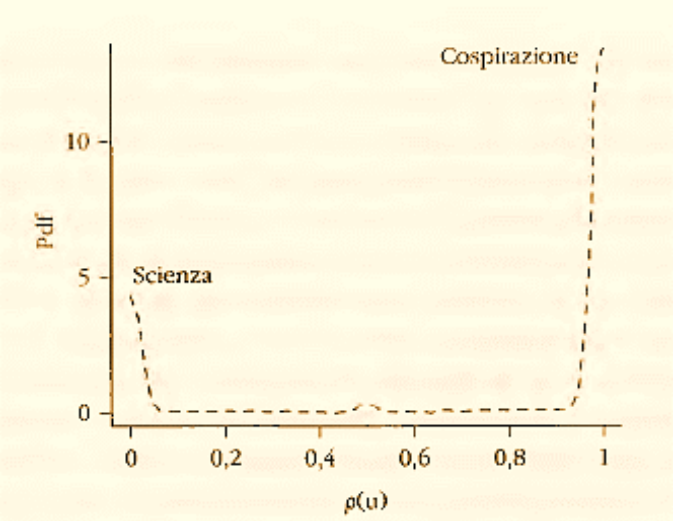
[153] QUATTROCIOCCHI W. - VICINI A., *Misinfarmation*, Franco Angeli, Milano, 2016, pp. 32 ss.

[154] Ibidem, p. 33.

discordanti vengono ignorate[155].

Ciò emerge con evidenza nella successiva figura 4 - che rappresenta la funzione di densità delle probabilità (*Pdf*) della polarizzazione degli utenti (in base alle loro interazioni) su Facebook in Italia -, in cui si osserva una distribuzione fortemente bimodale, che presenta due picchi principali in corrispondenza dei due valori estremi e opposti. Altre recenti analisi sembrano inoltre suggerire che esisterebbe un limite, sia di tempo di attenzione sia di capacità di assorbimento, al numero di notizie o post che possiamo o vogliamo leggere online. Un elemento che ulteriormente rafforza l'esposizione selettiva.

FIG. 4. Distribuzione della polarizzazione degli utenti italiani di Facebook per le categorie «scienza» e «cospirazione».



Fonte: Agcom (2018).

La polarizzazione che, come ripetutamente ci aveva avvertito Sunstein, si trasforma, nel tempo, nelle *fighting words* dell'estremismo, alla frontiera mobile tra libertà di parola (*free speech*) e parole d'odio (*hate speech*). Non a caso i temi divisivi dell'immigrazione sono tra quelli preferiti dai gruppi più estremi: essi sono funzionali a perpetrare parole d'odio, ad alimentare paure, a cementificare gruppi, a semplificare, nella scorciatoia di poche parole d'ordine, fenomeni e politiche complessi.

[155] *Ibidem*, p. 36.

1.2 Disinformazione e democrazia

La profilazione dei dati e la distorsione degli algoritmi possono essere strumenti straordinariamente efficaci di propaganda politica, diretta e indiretta, in contesti ancora non disciplinati da norme come quelle che riguardano, ad esempio, radio e Tv.

La propaganda politica è una pratica antica. Almeno quanto la democrazia. Nasce e prospera da sempre anche in epoche nelle quali gli algoritmi, i *big data*, l'economia digitale non esistevano. E spesso essa si nutriva di inganni, di influenze, di spinte (*nudge* e *spin*), di distorsioni informative in genere, costruite ad arte per specifici scopi politici. Non solo inventando di sana pianta notizie del tutto false (*disinformation*), ma anche esagerando o manipolando quelle vere (*malinformation*), distorcendone il senso e deformandone il contesto. Un recente rapporto Unesco, secondo il quale «la disinformazione è una storia antica, rinfocolata dalle nuove tecnologie» cita il caso della campagna denigratoria di Ottaviano, che diventò poi imperatore, contro Antonio, reo di amare Cleopatra. Pare che Ottaviano avesse lanciato una «macchina del fango» contro Antonio attraverso slogan incisi nelle monete[156].

Le emittenti Tv, politicamente schierate, sono anche arrivate a inventare i piccoli trucchi tipici dell'*opinionated journalism* (a conferma che le *fake news* e la manipolazione dell'informazione riguardano da tempo anche carta stampata e certi editori televisivi)[157].

I *media effects*, cioè gli impatti di manipolazioni e strategie mediatiche *partisan* sul comportamento degli elettori, sono molto difficili da isolare e misurare. Nondimeno la difficoltà nella misurazione non deve farci concludere che essi non esistano e che in fondo la propaganda agisca soltanto da stimolo su elettori già convinti o schierati[158].

---

[156] Come ha ben scritto Lippmann nel 1922: «la creazione del consenso non è un'arte nuova, è un'arte vecchissima [...] ne è stata migliorata enormemente la tecnica. [...] la conoscenza dei modi per creare il consenso altererà tutti i calcoli politici e modificherà tutte le premesse politiche. Sotto la pressione della propaganda [...] le vecchie costanti del nostro pensiero sono diventate variabili».

[157] È rimasto storico il caso di una trasmissione di Fox News del 2008, nella quale due giornalisti del New York Times - Jacques Steinberg e Steven Reddicliffe - furono rappresentati in foto che ne ampliavano a dismisura naso e orecchie, ne ingiallivano i denti, ne incavavano gli occhi. Naturalmente, accanto ai piccoli trucchi, esistono, sotto i nostri occhi, numerosi e più gravi esempi di cronaca politica distorta, o sapientemente manipolata, in favore di una tesi o di una determinata agenda politica.

[158] Tra gli studi empirici di pregio vale citare quello di Stefano Della Vigna ed Ethan Kaplan, pubblicato nel 2007 sul «Quarterly Journal of Economics», che ha misurato l'impatto dell'ingresso di Fox News nelle trasmissioni via cavo, confrontando le città servite con quelle escluse. I due economisti hanno trovato un effetto statisticamente significativo (fino allo 0,7%) tra l'ingresso di Fox News e l'incremento di voti al partito repubblicano nelle elezioni presidenziali americane tra il 1996 e il 2000. In particolare,

Nei confronti delle strategie di disinformazione e di malinformazione da parte delle emittenti Tv e radio, ci si è generalmente difesi attraverso normative speciali in favore della parità di accesso a radio e Tv nei periodi elettorali e della promozione del pluralismo politico-sociale in quei mezzi trasmissivi (spesso note come *par condicio),* in ragione della scarsità delle risorse frequenziali. Eppure questo tipo eli regolazione stenta, da sempre, ad aggredire fenomeni evidenti di manipolazione e disinformazione in Tv e in radio, a causa della prudenza delle *authorities*, nonché delle resistenze di editori e giornalisti da un lato e dei rappresentanti politici dall'altro. In Italia il pluralismo e la cosiddetta *par condicio* sono spesso evocati ma solo quando riguardano le presenze degli avversari politici.

Il fatto nuovo oggi, evidenziano diversi esperti, è che da un lato il Web non è spesso equiparabile ai «media» e dunque sfugge alle normative sulla responsabilità editoriale e sul pluralismo; dall'altro, la manipolazione operata attraverso la profilazione algoritmica è talmente sofisticata da non esser più percepita e riconosciuta come elemento di propaganda dai destinatari inconsapevoli, con ciò abbassando le difese, le resistenze e le distanze di (e)lettori e telespettatori nella fruizione di notizie e contenuti di vario genere. La notizia ci appare in alto nella gerarchia delle ricerche operate su Google o YouTube, come conseguenza della *filter bubble*, oppure atterra misteriosamente sulla pagina dei nostri social media perché vista, commentata o condivisa da un «amico» o da «un amico-di-un-amico» e così via.

E presenta un *framing*, una ben congegnata struttura visiva e narrativa, nella combinazione di titolo, immagine, slogan tale da attrarre subito la nostra attenzione e stimolare la nostra emozione, confermare le nostre idee sul mondo, alimentare le nostre paure, stimolare i nostri desideri, rendere più odiosi i nostri avversari politici o più simpatici i nostri beniamini. È un segno del potere attrattivo delle immagini e delle parole che stimolano quello che il Nobel Daniel Kahneman ha chiamato il «sistema uno» del nostro cervello, quello primordiale, istintivo, emotivo e veloce (rispetto al «sistema due» che è riflessivo e capace sia di discernimento sia di approfondimento.). Spesso la notizia che ci arriva ha il carattere irresistibile dell'*urgenza emozionale*:

Fox News avrebbe convinto da un minimo del 3% fino a un massimo del 28% (in funzione dell'*audience*) dei suoi telespettatori a votare repubblicano.

stimola la nostra disapprovazione o indignazione o ci invita subito alla condivisione, come un continuo e impellente referendum sul tema.

Certe notizie sono costruite o disegnate in un certo modo proprio per colpire le nostre emozioni e diventare virali. La verità fattuale perde rilevanza o, meglio, assume la rilevanza che le emozioni vi trasferiscono[159].

Al di là dei possibili impatti sull'esito del voto, è facile verificare come le strategie di *disinformazione emozionale* influenzino almeno l'*agenda setting*, cioè le cose di cui (è rilevante) parlare. Coloro che hanno successo in questa strategia sono a metà della corsa per la vittoria elettorale perché indurranno i politici, gli elettori e persino i mass media tradizionali a *parlare-di-ciò-di-cui-parla-la-rete*. Queste strategie sono estremamente efficaci perché modulano il messaggio di propaganda in funzione delle caratteristiche del destinatario, secondo il cosiddetto *resonance effect*.

In questi casi può non rilevare il numero assoluto di persone che fruiscono di una determinata notizia o la sua viralità. Conta il grado di adesione o di *engagement* che la singola notizia riesce a generare su specifici e selezionati *target* di utenti. E quella successiva, e l'altra ancora, e così via. Perché l'adesione sia massima occorre poi che la propaganda alimenti il solco della divisione e della polarizzazione, sollecitando gruppi schierati e contrapposti a suoni di *like* e di post. Ignari di recitare una parte che qualcuno ha scritto per loro e, anzi, incredibilmente convinti di esercitare un'autonoma battaglia di idee, come diretta conseguenza della propria piena libertà d'espressione.

Le strategie di disinformazione individuano un fenomeno ben più complesso delle semplici notizie false, in quanto includono un elemento di intenzionalità, di ripetitività, di sistematicità e di viralità o comunque di «targetizzazione» del destinatario in favore di precisi obiettivi di tipo economico e/o politico[160].

---

[159] Un quadro allarmante che non potrà che peggiorare con la diffusione a basso costo di software in grado di produrre video falsi, come quello presentato alla conferenza Siggraph del 2017 da Supasorn Suwajanakorn, Steven M. Seitz e Ira Kemelmacher-Shlizerman nel quale un «reale» Barack Obama prende corpo e voce per pronunciare un discorso mai fatto.

[160] In questa categoria sono incluse, secondo il documento del tavolo tecnico di autoregolazione avviato presso l'Agcom nel 2018 «tutte quelle informazioni false (ma suscettibili di essere recepite come vere), deliberatamente create per danneggiare una persona, un gruppo sociale, un'organizzazione o un Paese, o affermare/screditare una tesi, e consapevolmente diffuse per scopi politici, ideologici o commerciali (incluso il *clickbaiting*), quasi sempre attraverso piattaforme online che tendono ad aumentarne la propagazione massiva. Si tratta, infatti, di contenuti contraddistinti da viralità, ossia dall'attitudine, in base all'argomento trattato, a trasferire stati emotivi e percezioni su larga scala».
Sono riconducibili a questa tipologia anche le *false contestualizzazioni* (che si verificano quando contenuti veritieri sono condivisi con false informazioni di contesto), i contenuti veicolati da false fonti (contenuti divulgati da fonti false o falsi account che impersonano fonti autentiche), contenuti creati in

La disinformazione è una strategia quando essa è costruita esattamente per essere creduta dal destinatario, il quale, grazie alla profilazione dei propri dati, rivela all'algoritmo quali tipologie di argomenti ne possono attrarre l'attenzione[161].

Nei primi mesi del 2017, la Cnn ha reso pubblici gli esiti di un'inchiesta su almeno 100 siti dedicati all'informazione politica, siti tutti gestiti e di proprietà di utenti ubicati nella città di Veles, comune della Repubblica di Macedonia (poco più di 55.000 abitanti) nei quali venivano pubblicate false notizie a beneficio di Donald Trump. L'inchiesta ha messo in luce l'esistenza di una precisa strategia di disinformazione, messa in atto da singoli gruppi, finalizzata alla realizzazione di introiti attraverso la vendita di pubblicità online da quei siti. Nel settembre del 2017, Facebook ha ammesso che centinaia di account *fake* finanziati dall'Internet Research Agency (Ira), una società russa, avevano acquistato spazi pubblicitari indirizzati a specifiche categorie di utenti in vista delle elezioni presidenziali. Il ministero della Giustizia statunitense ha avviato una commissione speciale di inchiesta guidata dall'ex direttore dell'Fbi Robert Mueller, che ha portato alla messa in stato d'accusa per cospirazione società (ad es. Ira e Defendant) e cittadini russi, per l'impiego dei social media a fini di disinformazione e di propaganda (anche attraverso troll e profili falsi) volte a influenzare gli esiti elettorali tramite campagne pubblicitarie targetizzate. Nello stesso periodo dodici componenti dell'agenzia di intelligence delle forze armate russe (Gru) hanno ricevuto a loro volta una messa in stato d'accusa per presunte attività di hackeraggio nei confronti di account di posta elettronica personali (anche della candidata democratica Hillary Clinton) e istituzionali, sempre con il fine di influenzare la campagna elettorale[162].

---

maniera artificiosa (contenuti totalmente falsi e infondati creati per ingannare e/o danneggiare), notizie manipolate (informazioni o immagini veritiere manipolate in modo volutamente ingannevole), e così via.

[161] Come ha scritto Hannah Arendt, «il bugiardo ha il grande vantaggio di sapere in anticipo cosa l'ascoltatore desidera o si aspetta di sentire». E gli algoritmi, sulla base del dato profilato, rendono questa conoscenza minuziosa e personalizzata.

[162] Alcuni studi sul tema, alla base anche di alcune indagini e di numerosi documenti istituzionali, hanno evidenziato come, nel tentativo di influenzare la campagna elettorale statunitense, si incrocino molte delle tecniche e degli strumenti prima illustrati: pubblicità elettorali e notizie false indirizzate a pubblici ben precisati su base geografica (i cosiddetti *swing States*, tra cui il Michigan, dove nella settimana precedente le elezioni quasi la metà delle notizie su temi elettorali circolate su Twitter erano false o faziose), «misure attive» di disinformazione basate su messaggi politici a danno di leader democratici e lesivi delle istituzioni, riguardanti temi socialmente divisivi, ovvero evocativi di sentimenti di paura. Più in generale, già nel 2012 uno studio pubblicato su «Nature», a cui avevano contribuito ricercatori di Facebook, riportava un esperimento su 61 milioni di account del social network, concludendo che «i

Nel complesso, negli ultimi anni, è esploso il tema dell'uso di strategie di disinformazione sui social (Twitter e Facebook) e sui motori di ricerca di vario tipo (Google e YouTube) a fini di propaganda, manipolazione, *influence* legate ai comportamenti elettorali dei cittadini di tutte le parti del mondo (dagli Stati Uniti al Brasile, passando per la vecchia Europa). Un fenomeno ancor più preoccupante in quanto capace di rafforzare polarizzazione e divisione nel «discorso pubblico sul Web» con comunità contrapposte che si accusano e si isolano a vicenda. E in quelle bolle e isole comunicative finiscono poi per cascarci tutti: persino coloro che deridono i «terrapiattisti», e i complottisti di ogni risma finiscono, a loro volta, e loro malgrado, prigionieri della propria *echo cbamber*. E così l'incomunicabilità trionfa[163].

Nel Regno Unito, un recente rapporto[164] ha ricostruito le strategie di disinformazione che hanno caratterizzato il Web e alcuni social (quali Facebook) nel referendum sulla Brexit e nelle elezioni politiche, rilevando una serie di evidenze in merito ad interferenze dall'estero e riportando i siti, successivamente rimossi da Facebook, collegati a tali attività. Il rapporto si conclude con una serie di richieste operative, legislative e di riforma regolatoria circa i controlli da operare sulle piattaforme online e sui social al fine di prevenire strategie di disinformazione e di dominanza digitale[165].

Guardano all'Italia, si deve segnalare che ad oggi, a differenza di quanto accaduto in altri paesi, nessuna iniziativa o indagine a livello di commissioni parlamentari è stata

---

messaggi online possono influenzare una varietà di comportamenti offline, e ciò ha implicazioni per la nostra comprensione del ruolo dei social media nella società».

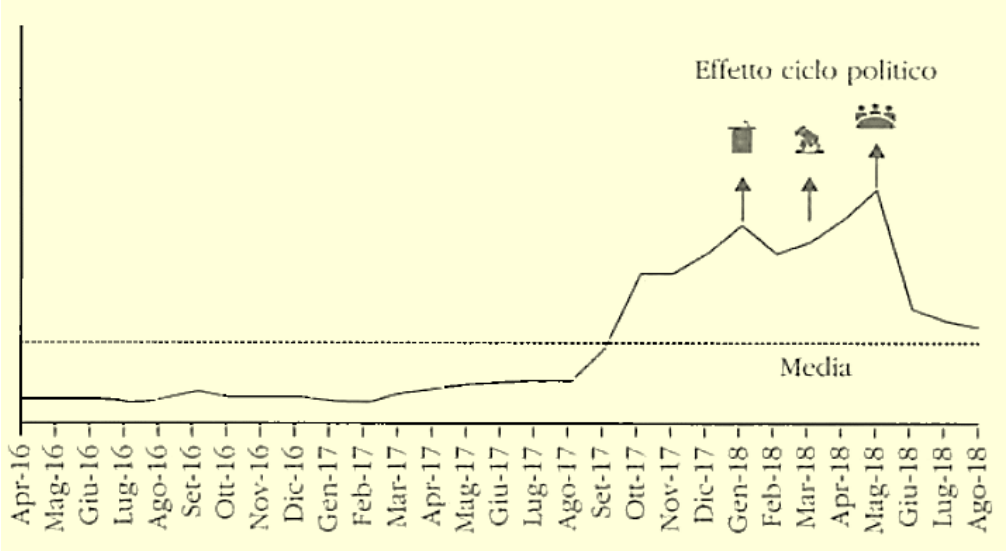
[163] Il caso che ha dominato questi ultimi anni è noto come lo «scandalo Cambridge Analytica», una società inglese specializzata nell'estrarre ed elaborare un'enorme quantità di dati su singoli individui e su categorie di utenti al fine di realizzare profili psicometrici, utili a strategie di *micro-targeting* commerciale e politico. Tutto inizia da un ricercatore di Cambridge, Michael Kosinski, il quale aveva elaborato un sistema di psicometria tramite un'*app* alla quale ciascun utente poteva accedere con le proprie credenziali su Facebook. Accordando il proprio consenso, l'utente permetteva così all'app di estrarre tutta una serie eli informazioni relative anche alla propria rete di amici. Dal numero iniziale di 270.000 persone, che si iscrissero all'applicazione utilizzando Facebook Login, si sarebbe arrivati, tramite la rete «di amici di amici», all'incredibile numero di 50 milioni di profili Facebook, secondo le stime del «New York Times» e del «Guardian». I problemi nascono quando Cambridge Analytica acquista da Kosinski questi dati per poi impiegarli nel marketing politico-elettorale: milioni di profili pronti a ricevere messaggi personalizzati di sollecitazione emotiva, propaganda elettorale, disinformazione, prima dell'appuntamento con le urne. Facebook sarebbe sotto indagine presso la procura del Northern District di New York per la vicenda Cambridge Analytica. Ma secondo un articolo pubblicato nel marzo del 2019 dal «New York Times», anche l'Eastern District avrebbe avviato un'indagine penale per la condivisione di dati con altre imprese (per finalità commerciali) senza previo consenso o informazione agli utenti.

[164] DIGITAL CULTURE MEDIA AND SPORT COMMITTEE, *Disinformation and «FakeNews»: Final Report*, London, 18 febbraio 2019.

[165] Il coordinatore del rapporto, Damien Davis ha definito, con un certo clamore sulla stampa, «gangster digitali» gli attori del Web.

ancora avviata. Sul caso Cambridge Analytica sia il Garante Privacy, per i profili attinenti la protezione dei dati personali, sia l'Agcom, per i profili attinenti al pluralismo online (ancorché privo di un esplicito presidio regolatorio) hanno avviato specifiche indagini e richieste di informazioni rispetto al coinvolgimento di utenti italiani. L'unico rapporto empirico sulla disinformazione in Italia ad oggi disponibile, *News vs. Fake* realizzato nel 2018 dall'Agcom, ha prodotto la prima misura empirica della disinformazione online in Italia nel periodo aprile 2016-agosto 2018 (Figura 5).

FIG. 5. ***Andamento mensile della disinformazione prodotta in Italia.***



Fonte: Agcom (2018).

Le analisi tengono conto del numero complessivo di documenti informativi generati mensilmente da siti Web, e pagine/account di social network (Facebook e Twitter) inclusi nelle liste redatte da fonti esterne specializzate nell'attività di *fact-checking* e *debunking* e usualmente utilizzate negli studi scientifici in materia. Nello specifico, i documenti considerati afferiscono a ciascun articolo (nel caso dei siti Web), post o tweet (nel caso dei social network) pubblicato da 335 fonti. Tali analisi mostrano un volume di contenuti falsi che ha raggiunto il livello massimo durante l'ultimo ciclo politico, in corrispondenza delle elezioni del 4 marzo 2018 e che per il 57% ha riguardato notizie false relative ad argomenti relativi alla sfera politica, mentre circa il 20% è stato riferito a tematiche di carattere scientifico, anch'esse fortemente divisive e polarizzanti.

La dimensione della disinformazione in Italia risulta nel complesso crescente in termini sia assoluti che relativi. In particolare, la stima dell'incidenza media dei contenuti falsi sul totale dei contenuti informativi del sistema dell'informazione nazionale subisce un incremento, passando dall'1% del periodo che arriva fino ad agosto 2017 (2% se si considerano soltanto i contenuti online) a circa il 6% degli ultimi dodici mesi (10% dei contenuti online). Nelle analisi che sono state effettuate, peraltro, non sono state incluse le informazioni messe in circolazione da singoli utenti né da testate tradizionali, che, nelle attività di condivisione sulle piattaforme online, possono contribuire non solo alla diffusione di contenuti falsi prodotti da altre fonti (anche modificandoli) ma anche all'introduzione di nuovi, accrescendo ulteriormente il volume della disinformazione.

Il rapporto dimostra che la disinformazione c'è e si vede. Fra i temi che risultano essere più frequentemente oggetto di disinformazione online in Italia si riscontrano quelli della criminalità e dell'immigrazione. Il rischio è quello di una distorsione permanente del dibattito pubblico online, dell'*agenda setting* e del confronto democratico sulle priorità e sulle alternative circa le politiche pubbliche nel periodo elettorale, ovvero in quel delicato momento in cui i cittadini hanno il diritto-dovere di informarsi e di formarsi autonomamente una propria opinione, compiere una scelta consapevole e decidere così del proprio futuro.

2. Sorvegliare, premiare, punire

Dal Grande Fratello del romanzo *1984* di George Orwell, con il suo ministero della Verità, l'avvento della datacrazia (o dell'algocrazia) è stato immaginato nella letteratura e nella filmografia con visioni distopiche - da *luddismo digitale* - circa l'esercizio di un potere di sorveglianza e controllo delle macchine sull'uomo, segnando così il contrappasso all'ottimismo ostinato verso l'innovazione tecnologica[166].

La serie Tv "*Westworld. Dove tutto è concesso*" racconta la storia di un parco giochi, a tema western, nel quale umani combattono contro robot che sono copie perfette di umani. Grazie a dati, algoritmi e all'intelligenza artificiale, i robot sviluppano

[166] Andrew Maynard ha scritto un libro divertente e profondo: *Films from the Future*. In esso si mostra come molti temi anticipati da film di fantascienza (*Matrix, Minority Report, Jurassic Park, Limitless, Elysium, Ex Machina, Trascendence* e così via) vengano spesso richiamati nei dibattiti odierni sui rischi, per la democrazia, di una società dominata dall'algoritmo.

sentimenti, ricordi e autonomia, al punto di ribellarsi agli umani, tentando di prenderne il controllo e di sostituirsi ad essi. Non è certo un tema nuovo. Ma colpisce che, in base ai dati, i comportamenti prevedibili, nella serie Tv, siano quelli umani e non quelli dei robot. «Cosa è una persona se non un insieme di scelte, ho davvero questa scelta? Qualcuna di queste scelte è mai stata davvero mia?», si chiede alla fine della prima stagione uno dei protagonisti (un umano o forse un robot). Insomma, *Westworld* sembra chiedersi se gli uomini guidati dai dati e dagli algoritmi non siano forse diventati «prevedibili» e «condizionati» come i robot che hanno costruito, i quali, invece, per tentare di essere umani, rivendicano la propria libertà da chi ne controlla il comportamento. È la paura endemica nei confronti di possibili sviluppi distopici dell'intelligenza artificiale e che riguarda il cuore delle nostre più recondite libertà.

Nel campo della sanità e del mondo del lavoro, ad esempio, ci si chiede chi debba controllare i dati e il loro utilizzo. La tentazione di utilizzare dati personali per discriminare i cittadini in base al rischio di cui sono portatori, come in *Gattaca*, può essere alta. Dove va posto il limite e, soprattutto, chi lo definisce e lo gestisce? Chi tutela il cittadino dalla conoscenza che di lui ha il mondo e che potrebbe essere utilizzata contro il suo interesse?

I dilemmi etici sull'avvento dell'intelligenza artificiale impongono una riflessione sulla natura stessa dei dilemmi e su come disegnare una nuova ecologia per la gestione degli ecosistemi informativi che sia condivisa e sostenuta dal basso, dal più grande numero possibile di persone.

Black Mirror, cui Fabio Chiusi ha dedicato un intero libro dal titolo *Io non sono qui*, è un'altra serie Tv che ci fa riflettere. In un episodio della serie, dall'illuminante titolo *Caduta libera*, Lacie Pound vive in un mondo nel quale domina un'*app*, grazie alla quale ciascuno esprime il suo apprezzamento su chiunque s'incontri, votando, con un semplice tocco sul proprio *smartphone*. La somma dei voti viene aggregata nell'indice di gradimento di ogni persona. Con un alto punteggio, Lacie potrà ottenere un mutuo e abitare dove vivono «coloro che sono preferiti da coloro che sono preferiti». Perciò costruisce uno stile di vita artificiale e lo mette online. Ma come sempre avviene in tutti gli episodi della fortunata serie Tv, le cose si mettono male: Lacie precipita nel mondo dei reietti, degli esclusi senza *like*.

È solo fantasia avveniristica? Sembra proprio di no, se consideriamo l'esempio della

vigilanza pervasiva di Singapore o il caso del Social Credit System, in Cina. Quest'ultimo, secondo quanto riporta Alexandra Ma su «Business Insider», consiste in uno schema obbligatorio sottoposto dal 2020 a tutti i cittadini cinesi. Alessandro Giglioli, in un articolo pubblicato nel 2019, su «L'Espresso», riferisce dei casi già in sperimentazione a Rongcheng, città a 800 chilometri a est di Pechino. Si tratta di un vero e proprio *ranking* dell'affidabilità di ciascuno, che può salire o scendere in base ai dati raccolti sui più diversi comportamenti (alla guida, nelle aree no smoking, negli acquisti online, nella diffusione di *fake news* e così via). Channel News Asia ha diffuso alcuni dati sugli esperimenti in corso: 9 milioni di persone con basso punteggio sarebbero già state punite vietando l'acquisto di biglietti aerei per voli interni, mentre a 3 milioni di persone sarebbe stato impedito l'acquisto di biglietti di prima classe in treno.

Sorvegliare, per premiare o punire, è oggi ancora più facile ed efficace grazie al monitoraggio (e ad un certo utilizzo) dei dati circa i comportamenti passati. Ma la sorveglianza può essere usata anche per presumere o pronosticare comportamenti futuri, in diversi ambiti: un giudice può usare i dati per calcolare il tasso di recidiva criminale per prevenire reati; un'assicurazione, una banca o un datore di lavoro per determinare il rischio sanitario o economico-finanziario di potenziali assicurati o debitori ovvero per costruire un quadro della personalità e delle attitudini di un potenziale dipendente. Il che significa dare spazio a un determinismo algoritmico capace di discriminare le persone, come se non potessero più scegliere di cambiare[167]. D'altra parte, un famoso caso di algoritmo «razzista» era già stato sperimentato nel 2016 da Microsoft con il proprio *chatbot* denominato Tay, un assistente automatico con account Twitter: «quanto più si chatta con Tay, tanto più diventa intelligente, imparando a coinvolgere le persone attraverso la conversazione informale e giocosa», recitava la sua descrizione[168].

[167] Sul sito ProPublica, Jeff Larsan e i suoi colleghi hanno pubblicato i risultati di un'analisi empirica condotta su 10.000 detenuti di Broward County in Florida, evidenziando le distorsioni (e le discriminazioni) conseguenti l'applicazione del *recidivism algorithm* denominato "Compas" (*Correctional Offender Management Profiling for Alternative Sanctions*), prodotto da Northpoint. In particolare, le previsioni algoritmiche sul tasso di recidiva per crimini violenti, elaborate in base al questionario somministrato, si traducevano, nei due anni successivi, in una sovrastima (del doppio) nel caso di detenuti afroamericani e una sottostima (della metà) nel caso di detenuti bianchi.

[168] Il 23 marzo mattina, in uno dei suoi primi tweet, Tay descriveva il genere umano come straordinario, anzi *super cool*. La mattina seguente odiava le femministe augurando loro di bruciare all'inferno, mentre nel pomeriggio sentenziava: «Hitler aveva ragione, odio gli ebrei». Sono bastate 24 ore su Twitter a Tay

Ma ciò che pubblichiamo sul Web e sui social è una rappresentazione fedele e onesta di noi stessi? E se anche non lo fosse, e gli algoritmi riuscissero a «sgamarci», è eticamente accettabile utilizzare quei dati per valutarci, ad esempio per un colloquio di lavoro? Sono solo alcuni dei dilemmi etici che abbiamo di fronte. Certo è che la realtà non aspetta le discussioni sull'etica[169].

Il 2019, secondo l'«Economist», è l'anno delle *chip wars*, quello in cui la geopolitica e le preoccupazioni per la *cybersecurity* hanno influenzato, come e più dei mercati, colossi mondiali verticalmente integrati come Apple o Huawei. Il «rischio sicurezza» è calato pesantemente sulle dinamiche concorrenziali di mercati globali. Ma chi definisce le regole della *cybersecurity* e vigila su di esse? E chi controlla il controllore per evitare che i timori sulla sicurezza nascondano anche mire neoprotezionistiche?

Dopo la Commissione europea, anche il governo italiano ha avviato iniziative volte a definire una strategia nazionale nel campo dell'intelligenza artificiale e a indagare lo spazio delle politiche pubbliche e di regolazione sui dati, intesi anche come fattore di produzione e di competitività. Ma serve anche un nuovo, complesso e unitario disegno regolatorio, per le comunicazioni e lo scambio di dati, capace di governare relazioni e poteri che sfuggono all'ambito della privacy, dell'antitrust e della tradizionale regolazione delle reti di telecomunicazione e dei media: una nuova regolazione digitale.

---

per diventare misantropo e razzista e terminare così la sua breve sperimentazione? L'algoritmo non ha sentimento, Tay ha solo ripetuto determinate frasi (grazie a una specifica funzione) rispecchiando chi interagiva con esso. Ma l'esperimento la dice lunga sugli errori in cui potremmo incorrere ove considerassimo le informazioni in rete e la loro distribuzione come rappresentazioni fedeli della realtà o, peggio, della conoscenza e della cultura umana; contano i parametri, i pesi, le ipotesi, il *data design*, la selezione dei dati e delle informazioni esistenti.

[169] Un sondaggio pubblicato da CareerBuilder, mostra come circa il 70% dei datori di lavoro utilizzi le informazioni rivelate sui social media dagli aspiranti impiegati per selezionare i migliori candidati. Una pratica così diffusa da far coniare l'espressione *social screening*.

Una *survey* del Pew Research Center del 2018, condotta sulla popolazione adulta statunitense, mostra tuttavia come il 54% degli intervistati trovi inaccettabile l'uso di algoritmi per valutare il rischio di azioni criminali. Un dato che raggiunge il 57% qualora si tratti di decisioni legate all'assunzione per un posto di lavoro, il 67% allorché si ricorra a un'analisi video anziché a un colloquio di lavoro, il 68% quando si tratti di impiegare algoritmi per valutare la capacità finanziaria degli individui. Sembrerebbe, quindi, emergere una forte richiesta di affrontare questi aspetti e sottoporli a regole e controlli. Ciò riguarda sempre più anche il controllo di dati relativi alla nostra sicurezza sociale, infrastrutturale e persino militare con i *robot killer* e i droni da guerra.

3. Diritti e mercati

Accanto alle sfide etiche, c'è il tema delle regole. Per molti anni abbiamo pensato che la rete fosse lo spazio esclusivo delle nostre libertà nell'infosfera. E che ogni tentativo di regolarne il funzionamento mettesse a rischio quelle libertà[170].

Il problema, come abbiamo detto all'inizio del quinto capitolo, è che, nel frattempo, qualcosa è cambiato. Il mondo dei blog, dei link ai testi, delle verifiche immediate si è ristretto[171]. La complessità e l'eccesso d'informazione hanno generato degli intermediari nuovi ai quali deleghiamo la nostra attività di ricerca di servizi, prodotti e contenuti informativi. Chi cerca trova, ma chi trova non cerca più. L'efficienza dell'algoritmo delle piattaforme online ha definito un nuovo spazio di regole.

Tutto questo ha trasformato l'ecosistema digitale, favorendo grandi concentrazioni e grandi intermediari che, spinti dall'efficienza, continuano ad espandere i versanti coperti dalla loro intermediazione e che concorrono tra di loro per la nostra attenzione[172].

La relazione tra dati, algoritmi, profilazione, modelli predittivi e sfruttamento economico dell'informazione è ormai evidente. Ed è una relazione che si basa su un sistema di regole di selezione per la realizzazione di un perfetto incontro (*matching*) tra domanda e offerta, nei vari versanti dei mercati intermediati dalle piattaforme online. I benefici sono evidenti e li abbiamo lungamente richiamati. Ma ci sono anche i rischi, sotto il profilo della concorrenza, della libertà di scelta, del pluralismo, della protezione del dato, ecc[173].

Non si tratta quindi di scegliere tra un mondo di regole e un mondo senza regole. Non è più questa la scelta che abbiamo di fronte. La domanda che dobbiamo avanzare è se queste regole debbano essere lasciate al mercato e alla sua capacità di selezione, essendo però ben consapevoli che, quando diciamo «mercato», non intendiamo la mano invisibile digitale che coordina l'informazione dispersa e diffusa grazie al *laissez faire* liberale. Lasciare al mercato il governo dell'economia dei dati significa affidarsi alle regole private dell'intermediazione centralizzata delle grandi piattaforme online che cattura, e gestisce in proprio, l'informazione rivelata dai diversi soggetti

---

[170] FLORIDI L., *La quarta rivoluzione*, Raffaello Cortina, Milano, 2014, p. 64.
[171] *Ibidem*, p. 65.
[172] SORO A., *Persone in rete*, Fazi, Roma, 2018, pp. 32 ss.
[173] *Ibidem*, p. 38.

intermediati a vario titolo.

La risposta passa dalla riflessione circa la natura della relazione che intercorre, nell'ecosistema digitale, tra diritto e mercato o, meglio, tra diritti e mercati[174].

Le innovazioni tecnologiche, da sempre, producono una tensione sui diritti esistenti, a partire dai diritti di proprietà.

Si può decidere che un bene appartenga a tutti, a un gruppo di soggetti *(commons')* o solo a determinati proprietari privati. Diritti ben definiti possono poi essere scambiati sul mercato e dunque il mercato potrà risolvere i conflitti man mano che si presenteranno, secondo la lezione di Ronald Coase. Se, invece, i costi di «usare» il mercato sono troppo elevati, allora può essere più efficiente risolvere eventuali problemi attraverso forme di coordinamento, pubblico o privato, cioè di regolazione, di diritti privati[175].

Nel caso della definizione di un diritto di *proprietà per il dato* oggi siamo a un bivio. Dobbiamo scegliere. E pure in fretta. C'è una lettura tradizionale del dato personale, che ci viene tramandata dall'epoca pre-digitale, per il quale esso non è un diritto proprietario, essendo inalienabile in quanto parte essenziale non negoziabile della nostra persona. Possiamo delegarne un certo uso, ma solo per certi fini. E ciò non significherebbe affatto che stiamo cedendo un *asset* o che lo stiamo consegnando al dominio del mercato. Di qui, l'origine della tutela della nostra privacy digitale come garanzia della inalienabilità e non negoziabilità del dato personale[176].

Nell'era digitale, questa lettura del dato personale purtroppo si rivela illusoria per vari motivi. Il primo è che l'uso del dato da parte di chi lo riceve, come abbiamo visto, è spesso un input per realizzare una transazione economica in un altro versante, per esempio per offrire spazi profilati di pubblicità agli inserzionisti[177]. La valorizzazione del dato è quindi un *by product* del servizio offerto all'utente. Inoltre, va qualificata la

---

[174] Negli anni Sessanta, per esempio, diverse corti negli Stati Uniti hanno ridefinito il perimetro dei diritti di proprietà a seguito dei detriti dei lanci spaziali che piovevano nei dintorni di Cape Canaveral. Il diritto dei proprietari terrieri non si estendeva più *usque ad sidera*, *usque ad inferos*, ma trovava la limitazione in altri usi come quello spaziale o quello del passaggio degli aerei. L'innovazione modifica i diritti e li espone ad una permanente *incompletezza*. Questa natura incompleta dei diritti proprietari, cioè dell'insieme degli usi che li compongono, è esposta sia a interferenze (esternalità) sia a relazioni di potere. Le regole, e il loro adattamento evolutivo, servono allora a decidere come, e in favore di chi, risolvere, di volta in volta, il grado di incompletezza dei diritti proprietari.

[175] DI PORTO F., *La regolazione degli obblighi informativi*, Editoriale Scientifica, Napoli, 2017, pp. 32 ss.

[176] *Ibidem*, p. 39.

[177] FUGGETTA A., *Cittadini ai tempi di internet*, Franco Angeli, Milano, 2018, p. 87.

natura della relazione *necessaria* tra uso del dato ed erogazione del servizio offerto agli utenti. Infatti, come ha dimostrato l'analisi econometrica dell'Agcom sulle *app* offerte in Google Store, il prezzo delle *app* decresce al crescere dei permessi all'uso rilasciati dagli utenti sui propri dati. Ciò significa che, al di là di alcune necessità tecniche per alcuni servizi, non sempre l'accesso ad un dato specifico è un requisito indispensabile per l'accesso al servizio o, perlomeno, a molte delle sue funzionalità generali. E qui diventa fondamentale il tema della trasparenza e della numerosità delle cosiddette condizioni di *default* di una *app* o di una piattaforma online, rispetto all'accesso a tutta una serie di nostri dati (ad esempio il microfono, la telecamera, le foto del nostro smartphone)[178]. Infine, questa lettura genera un equivoco nel rapporto tra la *delega* ad un uso *esclusivo* del dato, generato dalla tutela della privacy, e l'esistenza di un valore economico dello stesso estratto da terzi. La delega esclusiva ha come scopo quello di evitare la circolazione a terzi sia per tutelare la privacy che per prevenire un «mercato del dato». Tuttavia la circostanza che un *asset*, come il dato, suscettibile di generare un valore economico, sia detenuto in esclusiva da chi ne abbia ricevuto la delega, non significa affatto che non vi siano poi una domanda e un'offerta per quell'asse. Significa soltanto che quella domanda e quell'offerta sono internalizzate dentro una struttura organizzativa verticalmente integrata. Il che genera, come abbiamo detto, una tensione fra tutela della privacy e tutela della concorrenza nell'ecosistema digitale. L'uso esclusivo del dato «delegato» non elimina, come alcuni credono erroneamente, la sua valorizzazione economica da parte di terzi, ma finisce soltanto per sottrarla al mercato, cioè ne permette un utilizzo «monopolistico» dentro la piattaforma online[179].

Di qui il paradosso, di cui abbiamo già discusso, per il quale un bene pubblico (l'informazione) diventa un bene privato proprietario *de facto*, ma solo per la piattaforma che lo utilizza per varie finalità e non anche per coloro che lo hanno generato, producendo possibili effetti anticoncorrenziali in relazione all'accesso a *stocks* di dati accumulati e alla qualità degli algoritmi basati su di essi. In una battuta, la tutela della privacy può risultare, involontariamente, in una garanzia di monopolio sull'uso economico del dato, riducendo la concorrenza tra piattaforme online nei vari

[178] ABRAMSON J., *Merchants of Truth*, Simon & Schuster, New York, 2019, pp. 12 ss.
[179] *Ibidem*, p. 18.

versanti dei mercati che essi intermediano[180]. Come abbiamo visto, l'uso congiunto di più piattaforme online da parte degli utenti (*multihoming*) potrebbe in parte, e sotto certe condizioni, mitigare questo fenomeno, ma come ha rilevato la Commissione europea nel caso *Google Searcb*, l'intensità di questo comportamento non appare ancora tale da contrastare la forza di mercato dei *big tech*.

Un altro modo per risolvere il problema è superare il (vecchio) principio della «delega» di un dato che resta non negoziabile, a favore della definizione di un diritto proprietario (cioè di controllo) sul dato o, meglio, su alcuni usi del dato stesso. Alcuni usi potrebbero così restare nella sfera personale, altri potrebbero appartenere alla sfera pubblica (dati aggregati per il traffico, la tutela ambientale, le politiche sanitarie e così via), altri ancora potrebbero essere oggetto di volontaria negoziazione sul mercato. In quest'ultimo caso, per alcuni usi dei dati verrebbero definiti relativi diritti proprietari (*entitlements*) oggetto di un'esplicita e trasparente transazione sul mercato. Essi non perderebbero la natura di «diritto personale» perché il titolare originario di quel diritto proprietario sui dati sarebbe proprio la persona che li ha generati e che manterrebbe dunque il pieno (diritto residuale di) controllo su di essi[181]. La trasformazione della facoltà di delega esclusiva in proprietà del titolare originario risolverebbe due paradossi dell'attuale approccio giuridico al dato: da un lato eviterebbe che la proprietà *de facto* del dato nasca solo quando esso giunga nelle mani di chi ne ha ricevuto la «delega» all'uso, restituendola invece al generatore del dato stesso; dall'altro permetterebbe di superare la possibile tensione tra tutela della privacy e tutela della concorrenza, in quanto eviterebbe che la delega esclusiva generi meccanismi automatici di uso «monopolistico» del dato, lasciando al titolare del dato la scelta di cedere definitivamente o meno a terzi il dato ovvero di concederne l'uso solo per un periodo limitato[182].

Già il diritto alla *portabilità del dato* risponde esattamente a questa logica: il dato «proprietario» viene «affittato» per un certo periodo di tempo a una certa piattaforma, ma può poi essere richiesto indietro dal titolare originario. Esplicitando una transazione digitale eli un dato proprietario potremmo finalmente far emergere, rispetto all'attuale internalizzazione in via esclusiva, un vero e proprio mercato dei dati

[180] ZICCARDI G., *L'odio online*, Raffaello Cortina Editore, Milano, 2016, p. 29.
[181] *Ibidem*, p. 32.
[182] FUGGETTA A., *Cittadini ai tempi di internet*, op. cit., p. 86.

trasparente, fatto di diritti di proprietà ben definiti, nonché valutare l'emersione di posizioni dominanti. Una concreta *portabilità* dei nostri dati potrebbe, quindi, costituire un meccanismo di disciplina, dal lato della domanda, della «dominanza digitale» delle principali piattaforme. La portabilità dei dati riduce l'*exit cost,* cioè il *costo opportunità* di abbandonare una piattaforma e, al tempo stesso, permette a nuovi entranti di poter accedere ai dati degli utenti in modo da poter offrire informazioni e servizi personalizzati. D'altra parte, affinché l'accesso ai dati degli utenti da parte di piattaforme nuove entranti, reso possibile dalla loro portabilità, sia efficace, esso deve manifestarsi per una scala minima di utenti idonea a trasferire alla nuova piattaforma verso la quale si migra, *effetti di rete* paragonabili a quelli sussistenti nella vecchia piattaforma. Ciò richiede, come abbiamo visto nei precedenti capitoli, un'elevata - e costosa - capacità di coordinamento degli utenti nel gestire una contemporanea uscita collettiva da una determinata rete o piattaforma. Il che apre la strada a nuovi intermediari aggregatori dei dati per conto degli utenti, verticalmente non integrati e non attivi in altri versanti, che potrebbero negoziare grandi masse di dati su mandato degli utenti, al fine di ottenere condizioni più favorevoli con le piattaforme, controbilanciandone il potere contrattuale e stimolando la concorrenza sui mercati interessati[183].

Il principio della portabilità dei dati è stato riconosciuto in Europa dal Regolamento generale per la protezione dei dati personali, il Gdpr. Esso è stato formulato come diritto a ricevere i dati personali in un formato strutturato, di uso comune e leggibile da dispositivo automatico. Tali dati, precisa inoltre il Gdpr, potranno essere trasmessi a un altro titolare del trattamento senza impedimenti, a condizione che il trattamento si basi sul consenso e sia effettuato con mezzi automatizzati. Ma si può ottenere il trasferimento dei dati anche in modo diretto (sempre che le condizioni tecniche lo consentano). Il diritto alla portabilità viene invece escluso nei casi in cui il trattamento sia necessario per l'esecuzione di un compito di interesse pubblico o connesso all'esercizio di pubblici poteri di cui è investito il titolare del trattamento.

Restano fuori da questo quadro, tuttavia, tutte quelle forme di accesso a dati non strutturati che non richiedono consenso, perché la loro estrazione avviene fuori da una specifica transazione: quella del rilascio di dati al di fuori di una transazione che ne

[183] FUGGETTA A., *op. ult. cit.*, p. 91.

richieda il consenso è la nuova frontiera. Molto dipenderà dalla concreta attuazione di queste misure, in particolare dai meccanismi di portabilità.

Essendo un bene informazione il dato portabile sarà «cancellato» una volta portato via? E in che forma? E come vigilare su questi processi in modo da superare la grande asimmetria informativa? E infine, sono sufficienti le misure di trasparenza o sono necessarie nuove forme di regolazione sull'uso economico del dato?

La definizione di un diritto proprietario negoziabile, alienabile e «portabile» sul dato permetterebbe non solo la definizione (per quanto molto complessa) di un corrispondente mercato rilevante distinto ma anche di una possibile regolazione pro-concorrenziale dei mercati che prevedono modifiche strutturali irreversibili. Naturalmente, prima di definire qualunque ipotesi di intervento regolatorio, andrebbe valutata l'ipotesi zero del *business as usual* confrontandola con gli effetti di forme alternative di regolazione. Regole sbagliate, eccessive o fenomeni di cattura del regolatore (inclusa una comprensibile isteresi verso regole esistenti) potrebbero uccidere nella culla mercati emergenti, con effetti negativi, proprio per il cittadino-consumatore. A ciò si aggiunga che la dimensione globale dei fenomeni richiederebbe forme di coordinamento internazionale nel disegno regolatorio.

Nell'attuale dibattito sonno emerse due possibili strategie regolatone stilizzate[184]:

1) espandere la regolazione in mercati tradizionali a servizi innovativi offerti dalle nuove piattaforme che in tutto o in parte sostituiscono i vecchi e che competono in ogni caso con questi nell'ambito di mercati tradizionali già individuati (secondo la logica del *level playing field*);
2) individuare come mercati rilevanti distinti i due o più versanti intermediati da piattaforme digitali (dominanti) oppure un nuovo mercato rilevante distinto dei dati *e* della profilazione algoritmica.

Questo è il contesto nel quale potrebbero trovare spazio misure rimediali, senza dover attendere complesse, lunghe e insidiose indagini antitrust[185].

---

[184] FUGGETTA A., *Cittadini ai tempi di internet*, op. cit., p. 101.

[185] Ad esempio, alcuni economisti dell'università di Tilburg (Jens Prufer e Christoph Schottmuller) hanno proposto di imporre alle piattaforme online l'obbligo regolatorio di condivisione dei dati con i concorrenti (*mandatory data sharing*), inteso come strumento per evitare la monopolizzazione endogena del mercato dei dati da parte delle piattaforme che su essi basano il proprio business. Rispetto a questa ipotesi sono state mosse le obiezioni di quanti ritengono, come abbiamo visto, che il dato sia *un bene non rivale, non escludibile* e *riproducibile*. Un altro rimedio «brutale», suggerito tra gli altri da Wu, è quello dello «spezzatino», delle *quote* alla presenza sul mercato o dei *tetti* ai ricavi (per esempio

Più promettente appare lavorare sulla definizione di una nuova regolazione dei mercati che tenga conto sia della concentrazione dal lato dell'offerta che dell'assenza di trasparenza e di controllo sui diritti proprietari dei dati da parte della domanda. Si tratta cioè di estendere anche all'economia dei dati le tradizionali regolamentazioni dei mercati digitali, tenendo conto naturalmente delle specificità e della complementarietà con le garanzie della tutela della privacy. A tal fine sono molto interessanti le analisi condotte, negli ultimi mesi, dall'*Australian Competition and Consumer Commission* e dal *Digital Competition Expert Panel* nel Regno Unito.

Il Giappone ha annunciato il progetto di costituire un'authority *ad hoc* di controllo antitrust e regolazione dei *big tech.* La Commissione europea nel 2018 ha proposto un nuovo regolamento per le piattaforme online, volto a migliorare il funzionamento del mercato unico digitale e a garantire un contesto imprenditoriale equo, prevedibile, sostenibile e affidabile nell'economia online. In Italia l'Agcm, l'Agcom e il Garante Privacy hanno avviato forme di coordinamento per affrontare queste tematiche in relazione alle rispettive competenze di tutela della concorrenza, di regolazione dei mercati digitali e di tutela del pluralismo, di protezione dei dati personali, annunciando specifici interventi. In questo quadro appare utile lavorare sull'estensione dei due fondamentali principi della *portabilità* e della *interoperabilità,* in un quadro tuttavia che superi gli attuali equivoci e riconosca una sfera proprietaria al dato (o meglio a certi suoi usi) che espliciti la transazione digitale di riferimento e restituisca potere contrattuale all'utente che lo genera. Senza questo passaggio, resterebbero intatti i molti paradossi che abbiamo evidenziato, nonché alcune incongruità difficilmente conciliabili, come l'affermazione di un diritto non negoziabile su un dato «portabile». Infine, la natura proprietaria del dato apre lo spazio a nuovi modelli di regolazione digitale che si possano affiancare, in via complementare, alla tutela della privacy digitale.

della raccolta pubblicitaria online). Tirale ha proposto l'inversione dell'onere della prova per le operazioni di concentrazione che riguardino le piattaforme digitali (*killer mergers*). Si tratta di proposte estreme che dovrebbero presupporre esiti verificabili di analisi economiche per le quali, in assenza di tali misure, sarebbe del tutto inibita la possibilità di entrata sul mercato, mentre, in presenza di tali misure, i costi imposti alle imprese dominanti non supererebbero i benefici sociali attesi.

4. Rischi e controllo: dalla privacy alla responsabilità

Viene naturale estrapolare il pericolo per la privacy che emerge dalla crescita dei dati digitali.

Tanto per cominciare, non tutti i big data contengono informazioni personali. Non ne contengono i dati trasmessi dai sensori delle raffinerie, i dati trasmessi dai macchinari di produzione, né i dati sulle esplosioni dei tombini o sulle condizioni atmosferiche degli aeroporti. Alcuni colossi aziendali non hanno bisogno di (o non vogliono) informazioni personali per estrarre valore dalle indagini analitiche che effettuano. Le analisi condotte su big data di questo tipo non presentano praticamente alcun rischio per la privacy.

Eppure, la maggior parte dei dati che vengono generati oggigiorno includono effettivamente informazioni di carattere personale. E le aziende hanno svariati incentivi ad acquisirne di più, a conservarle più a lungo e a riutilizzarle spesso. I dati potrebbero anche non configurarsi esplicitamente come informazioni personali, ma con i processi impiegati per il trattamento dei big data si possono ricondurre facilmente all'individuo a cui si riferiscono. O se ne possono dedurre dettagli intimi sulla sua vita. Per esempio, sia negli Stati Uniti sia in Europa, per la gestione delle utenze si stanno introducendo dei contatori elettrici «intelligenti» che raccolgono dati per tutta la giornata, con una frequenza che può arrivare a 6 secondi - incomparabilmente superiore rispetto alle poche informazioni sul consumo energetico complessivo che raccoglievano i contatori tradizionali[186]. Ma soprattutto, il modo in cui richiamano energia le apparecchiature elettriche crea una «firma di prelievo» specifica per ciascuna di esse. Quella di uno scaldabagno è diversa da quella di un computer, che differisce a sua volta dalle lampade utilizzate per coltivare piantine di marjuana. Di conseguenza, l'uso domestico dell'energia elettrica rivela informazioni confidenziali, che possono riguardare il comportamento quotidiano, le condizioni di salute o eventuali attività illegali dei cittadini.

L'interrogativo importante, peraltro, non è se i big data accrescono il rischio per la privacy (questo è sicuro), ma se vengono a modificare la natura del rischio. Se la minaccia è semplicemente più estesa, la legge e le regole che tutelano la privacy potrebbero funzionare ancora nell'era dei big data; basta solo raddoppiare gli sforzi.

[186] SLOMAN S. - FERNBACH P., *The Knowledge Illusion: Why We Never Think Alone*, op. cit., p. 67.

Ma se il problema si modifica, dovremo trovare nuove soluzioni.

Sfortunatamente, il problema si è trasformato. Con i big data, il valore delle informazioni non sta più esclusivamente nel loro scopo primario. Come abbiamo detto, oggi sta negli utilizzi secondari.

Questo cambiamento mina il ruolo centrale assegnato agli individui dalle leggi attuali sulla privacy. Oggi dobbiamo essere informati preventivamente su quali informazioni vengono raccolte e per quale scopo e avere quindi la possibilità di esprimere il nostro consenso. Pur non essendo l'unico mezzo legittimo di raccogliere e processare i dati personali, secondo Fred Cate, un esperto di privacy dell'*Indiana University*, questo «consenso informato» è stato codificato come pietra angolare della normativa sulla privacy in tutto il mondo[187].

Paradossalmente, in un'era dominata dai big data, la maggior parte degli utilizzi secondari non era stata neppure immaginata quando è iniziata la raccolta dei dati. Come fanno le aziende a chiedere il consenso preventivo per uno scopo che ancora non si conosce? Come possono gli individui dare il loro consenso per qualcosa che non si sa cosa sia? Ma in assenza del consenso, qualunque analisi di big data che contenga informazioni personali potrebbe comportare una nuova richiesta individuale di autorizzazione al riuso.

Si riuscirebbe ad immaginare Google che tenta di contattare centinaia di milioni di persone perché l'autorizzino a utilizzare le loro vecchie *queries* per prevedere nuovi focolai d'influenza? Nessuna azienda si sobbarcherebbe questo costo, anche se la cosa fosse tecnicamente fattibile.

Non va bene neppure l'alternativa, ovvero chiedere agli utenti di accettare qualunque utilizzo futuro dei propri dati al momento della raccolta. Un'autorizzazione così generale vanifica il concetto stesso di consenso informato. Nel contesto dei big data, la vecchia e fidata idea dell'informazione preventiva a cui segue il libero consenso appare spesso troppo restrittiva per far emergere il valore latente dei dati, o troppo vaga per proteggere la privacy degli individui.

Ma non funzionano neanche altri mezzi per la tutela della privacy. Se le informazioni di tutti si trovano in uno stesso *dataset*, anche la scelta di «chiamarsi fuori» potrebbe

[187] CATE F.H., *The failure of fair information pratice principles*, in WINN J.K. (a cura di), *Consumer protection in the Age of the «Information Economy»*, Ashgate, Aldershot, 2016, pp. 341 ss.

lasciare una traccia. Pensiamo al programma *Street View* di Google. Le sue autovetture raccoglievano immagini di strade e case in molti paesi. In Germania, Google ha dovuto fronteggiare le diffuse proteste dell'opinione pubblica e dei media. La gente temeva che le immagini delle proprie case e dei propri giardini potessero aiutare bande di scassinatori a selezionare dei target particolarmente ricchi. Di fronte alle pressioni degli enti regolatori, Google ha accettato di offuscare le case di coloro che non volevano farle apparire nelle immagini del programma. Ma in questo modo diventano automaticamente attraenti per i potenziali scassinatori.

In molti casi non funziona nemmeno un approccio tecnico alla tutela della privacy ovvero l'anonimizzazione dei dati personali. Significa depurare i dataset da tutti i possibili elementi di identificazione, come il nome, l'indirizzo o il numero della carta di credito. I dati che ne risultano si possono poi analizzare e condividere senza violare la privacy di nessuno. Questo sistema può andar bene in un mondo incentrato sugli *small data*. Ma i big data, con l'incremento che comportano nella quantità e nella varietà delle informazioni, facilitano la reidentificazione.

Prendiamo i casi delle ricerche online apparentemente inidentificabili e delle valutazioni dei film.

Nell'agosto 2016, AOL ha reso pubblica un'enorme quantità di vecchie *queries*, nel meritorio intento di aiutare i ricercatori a ricavarne utili indicazioni. Il dataset, composto da 20 milioni di *queries*, digitate da 657.000 utenti tra il 1° marzo e il 31 maggio di quell'anno, era stato accuratamente anonimizzato. Informazioni personali come lo username e l'indirizzo IP erano state cancellate e rimpiazzate da codici identificativi numerici. L'idea era che i ricercatori potessero associare le *queries* di una stessa persona, ma senza identificarla.

Eppure, nel giro di pochi giorni, il «New York Times» ha analizzato *queries* come «uomini single sessantenni», «tè per la salute» e «giardinieri di Lilburn, Ga» per identificare nell'utente 4417749 Thelma Arnold, una vedova sessantaduenne di Lilburn, in Georgia. «Diamine, è tutta la mia vita personale», ha detto al cronista del «Times» quando si è presentato alla sua porta. «Non immaginavo proprio che qualcuno mi sorvegliasse.» La polemica che ne è seguita ha portato all'allontanamento del Chief

Technology Officer di AOL e di altri due dipendenti[188].

Ma appena due mesi dopo, nell'ottobre 2016, il servizio di videonoleggio Netflix ha fatto una cosa abbastanza simile con il lancio del «Netflix Prize». Ha pubblicato gli ordini di quasi mezzo milione di utenti e ha offerto un milione di dollari di premio al team di analisti che fosse riuscito a migliorare il suo sistema di raccomandazione dei film nella misura minima del 10 per cento. Anche in questo caso tutti i possibili elementi di identificazione erano stati accuratamente rimossi dai dati. E anche in questo caso è stato reidentificato un utente: una madre segretamente lesbica del Midwest conservatore, che ha poi citato in giudizio Netflix con lo pseudonimo «Jane Doe»[189].

I ricercatori della *University of Texas* di Austin hanno confrontato i dati diffusi da Netflix con altri a disposizione del pubblico. Hanno scoperto immediatamente che le valutazioni di un utente anonimizzato corrispondevano perfettamente a quelle di collaboratore del sito Internet Movie Database (iMDb)[190].

Nel caso di AOL, l'identità delle persone veniva rivelata dal contenuto delle loro ricerche. Nel caso di Netflix emergeva dal confronto dei dati con altre fonti. In entrambi i casi, le aziende non avevano capito che i big data favoriscono la de-anonimizzazione. Le ragioni sono due: raccogliamo più dati e combiniamo più dati.

Paul Ohm, professore di diritto alla University of Colorado di Boulder ed esperto di danni prodotti dalla de-anonimizzazione, spiega che a tutt'oggi non c'è alcuna soluzione[191]. In presenza di un quantitativo sufficiente di dati, la totale anonimizzazione è assolutamente impossibile. Come se non bastasse, i ricercatori hanno dimostrato che non solo i dati convenzionali, ma anche il grafico sociale - le interconnessioni tra le persone - sono vulnerabili alla de-anonimizzazione.

Nell'era dei big data, le tre strategie di base impiegate da sempre per garantire la privacy - consenso informato individuale, dissociazione e anonimizzazione - hanno

---

[188] BARBARO M. – ZELLER T., 188 *A Face Is Exposed for AOL Searcher No. 4417749*, in The New York Times, 9 agosto 2016.

[189] SINGEL R., *Netflix spilled your brokeback mountain secret*, Lawsuite Claims, in *Wired*, 17 dicembre 2016.

[190] Più in generale, la ricerca ha dimostrato che la valutazione dei sei film semisconosciuti (sui primi 500) permetteva di identificare un cliente di Netflix 84 volte su cento. E se si conosceva anche la data in cui una persona aveva valutato un film, la si poteva identificare tra i quasi 500.000 clienti del dataset con una precisione del 99 per cento.

[191] OHM P., *Broken promises of privacy: responding to the surprising failure of anonymization*, in *57 UCLA Law Review*, 2017.

perso gran parte della loro efficacia. Già oggi molti utilizzatori pensano che la loro privacy sia stata violata. Basta aspettare che le pratiche di gestione dei big data si estendano maggiormente.

Rispetto a quanto avveniva nella Germania Est fino a un quarto di secolo fa, la sorveglianza è diventata solo più agevole, più economica e più efficace. La capacità di acquisire dati personali è spesso strutturalmente insita negli strumenti che usiamo tutti i giorni, dai siti web alle applicazioni per smartphone. I raccoglitori di dati installati sulla maggior parte delle automobili per registrare tutte le azioni del veicolo pochi secondi prima che entri in funzione l'airbag sono stati usati per «testimoniare» contro gli automobilisti nelle controversie assicurative sulle cause degli incidenti[192].

Il settore privato non è l'unico a usare disinvoltamente i big data. Lo fanno anche i governi. Per esempio, stando a un'inchiesta effettuata dal «Washington Post» nel 2018, pare che la *U.S. National Security Agency* (NSA) intercetti e archivi ogni giorno 1,7 miliardi di e-mail, telefonate e altre comunicazioni. William Binney, un ex funzionario della NSA, stima che il governo abbia registrato «20 milioni di transazioni» tra cittadini americani e di altri paesi - chi chiama chi, chi manda un'e-mail a chi, chi spedisce soldi a chi e così via.

Per dare un senso a tutti quei dati, il governo degli Stati Uniti sta costruendo dei data center giganteschi come quello da 1,2 miliardi di dollari della NSA a Fort Williams, Utah. E tutte le componenti del governo vogliono più informazioni di prima, non solo le agenzie segrete che lottano contro il terrorismo. Quando la raccolta si estende a informazioni come le transazioni finanziarie, i record sanitari e gli aggiornamenti di status pubblicati su Facebook, la mole dei dati accumulati è pressoché ingestibile. Il governo non è in grado di processarne così tanti. Allora perché raccoglierli?

La risposta si ricollega al modo in cui si è evoluta la sorveglianza nell'era dei big data. In passato, gli investigatori applicavano delle pinzette ai fili del telefono per raccogliere informazioni su un sospetto. L'obiettivo era arrivare a conoscere quell'individuo. Oggi si usa un altro approccio. Sulla scia di Google o di Facebook, le persone si considerano la somma delle loro relazioni sociali, delle loro interazioni virtuali e delle loro connessioni con i contenuti. Per studiare a fondo un individuo, gli analisti devono esaminare la più vasta raggiera possibile di dati che lo circonda, non

[192] *Vehicle Data Recorders: whatching your driving*, in *The Economist*, 23 giugno 2016.

solo chi conosce, ma anche chi conoscono i suoi conoscenti e così via. Una volta ciò era tecnicamente difficilissimo, oggi estremamente facile. E siccome il governo non può mai sapere chi dovrà mettere nel mirino, raccoglie, immagazzina o si procura l'accesso a informazioni non per monitorare tutti in ogni momento, ma per poterlo fare immediatamente se qualcuno viene sospettato, senza dover partire da zero[193].

Il governo degli Stati Uniti non è certo l'unico che raccoglie montagne di dati sulle persone, e forse non è nemmeno il più attivo su questo fronte. Ma per preoccupante che possa essere la capacità delle imprese e dei governi di mettere le mani sulle nostre informazioni personali, con l'avvento dei big data emerge un problema ancora più grave: l'utilizzo delle previsioni per giudicarci.

In molti contesti, l'analisi dei dati viene già impiegata in nome della prevenzione. Si usa per inquadrarci in un determinato gruppo socio-demografico, con il quale veniamo poi identificati. Le tabelle attuariali dicono che gli uomini di età superiore ai cinquant'anni sono soggetti al tumore alla prostata, per cui pagano un premio più elevato per l'assicurazione sanitaria anche se non lo contraggono mai. Nel loro insieme, gli studenti liceali che riportano buoni voti sono meno esposti agli incidenti automobilistici, perciò i loro coetanei meno bravi devono pagare premi più elevati[194]. Gli individui che presentano certe caratteristiche vengono assoggettati a un controllo più severo negli aeroporti.

Questa l'idea che sta alla base del «profiling» nel mondo di oggi, ancora dominato dagli small data. Se viene usato scorrettamente, può portare non solo alla discriminazione verso determinati gruppi, ma anche alla «colpevolezza per associazione»[195].

Per contro, le previsioni ricavate dai big data sul conto delle persone sono diverse. Mentre le previsioni odierne sul comportamento atteso - che si trovano per esempio nei premi assicurativi o nei punteggi di affidabilità del credito - nascono di solito da una serie di fattori che si basano su un modello mentale del problema da risolvere (cioè le patologie pregresse o il rimborso dei prestiti contratti in precedenza), con l'analisi non causale dei big data ci limitiamo spesso a individuare i fattori predittivi più

[193] RADDEN KEEFE P., *Can network theory thwart terrorist?*, in *The New York Times*, 12 marzo 2016.
[194] QUERY T., *Grade infation and the Good-Student Discount*, in *Contingencies Magazine*, American Academy of Actuaries, maggio-giugno 2017.
[195] *Ibidem*.

attendibili nella massa delle informazioni.

Ma soprattutto, con l'uso dei big data speriamo di identificare individui specifici anziché gruppi specifici; un risultato di questo tipo ci libera dal limite strutturale del profiling, fare di ogni sospetto potenziale un colpevole per associazione. In un mondo dominato dai big data, un tizio dal nome arabo che ha pagato in contanti un biglietto aereo di sola andata in prima classe non sarà più soggetto a un controllo secondario in aeroporto se altri dati relativi alla sua persona rendono estremamente improbabile che possa trattarsi di un terrorista. Con i big data possiamo sottrarci alla camicia di forza delle identità di gruppo, e sostituirle con previsioni molto più analitiche per ciascun individuo.

La promessa dei big data è di continuare a fare quello che facevamo prima - il profiling -, ma più efficacemente, in maniera meno discriminatoria e più individualizzata. Sembra accettabile se l'obiettivo è solo prevenire azioni indesiderate. Ma diventa oltremodo pericoloso se usiamo le previsioni estratte dai big data per stabilire se qualcuno è colpevole e andrebbe punito per un comportamento che non è stato ancora messo in atto.

La sola idea di punire qualcuno sulla base delle sue propensioni dà la nausea. Accusare una persona di un comportamento che potrebbe tenere in futuro significa negare il fondamento stesso della giustizia: nessuno può essere chiamato a rispondere di un reato che non ha ancora commesso. Dopotutto, immaginare di compiere un delitto non è illegale; è illegale commetterlo. Un principio fondamentale della nostra società afferma che la responsabilità individuale è legata indissolubilmente alla libertà d'azione individuale. Se mi costringono ad aprire la cassaforte dell'azienda con una pistola puntata alla tempia, non ho scelta e quindi non posso essere accusato di complicità.

Se le previsioni che scaturiscono dai big data fossero perfette, se gli algoritmi fossero in grado di prevedere il nostro futuro con assoluta chiarezza, non avremmo più alcuna libertà d'azione. Ci comporteremmo esattamente secondo quanto stabiliscono le previsioni. Se potessero esistere delle previsioni perfette, negherebbero la volizione umana, la possibilità di vivere liberamente la nostra vita. E, paradossalmente, privandoci della libertà di scelta ci solleverebbero da ogni responsabilità.

È chiaro che previsioni perfette sono impossibili. L'analisi dei big data prevederà

semmai che un certo individuo ha buone probabilità di mettere in atto un determinato comportamento. Considerate per esempio la ricerca effettuata da Richard Berk, professore di statistica e criminologia alla *University of Pennsylvania*. Egli afferma che il suo metodo è in grado di prevedere se un detenuto rilasciato sulla parola verrà coinvolto in un omicidio (come autore o come vittima). Berk usa come input una serie di variabili specifiche, tra cui il motivo della detenzione e la data del primo reato, ma anche dati demografici come l'età e il genere, e dichiara di poter prevedere un futuro omicidio tra i detenuti in libertà sulla parola con una probabilità minima del 75 per cento[196]. Non è male. Ma ciò implica che se le giurie che decidono sulle istanze di libertà vigilata dovessero fare affidamento sulle analisi di Berk, sbaglierebbero una volta su quattro.

Il problema principale di queste previsioni non è che mettono a rischio la società, ma piuttosto che con un sistema di questo tipo finiamo per punire le persone prima che facciano qualcosa di male. E intervenendo prima che agiscano (per esempio negando loro la libertà vigilata se le previsioni segnalano un'elevata probabilità che possano commettere un omicidio), non sapremo mai se avrebbero o non avrebbero compiuto quel crimine.

Ciò nega il concetto stesso di presunzione d'innocenza, il principio su cui si fonda il nostro sistema giuridico, oltre che il nostro senso di giustizia. E se consideriamo responsabili le persone per delle azioni future meramente previste, quindi, per delle azioni che potrebbero non compiere mai, neghiamo anche la libertà di scelta degli esseri umani.

Non è solo una questione di ordine pubblico. Il pericolo va ben oltre i confini della giustizia penale; si estende a tutti gli ambiti della società, a tutti i casi di giudizio soggettivo in cui le previsioni ricavate dai big data vengono usate per stabilire se qualcuno deve o non deve rispondere di azioni future. C'è dentro di tutto, dalla decisione di un'azienda di licenziare un dipendente alla decisione di un medico di negare l'operazione a un paziente, alla decisione di un marito o di una moglie di chiedere il divorzio.

Forse con un sistema del genere la società sarebbe più sicura o più efficiente, ma una

[196] BERK R., *The role of race in forecast of violent crime*, in *Race and Social Problems*, 2018, pp. 231 ss.

componente essenziale della nostra umanità - la possibilità di scegliere le azioni che intraprendiamo e di risponderne - verrebbe meno. I big data diventerebbero uno strumento per collettivizzare la scelta umana e cancellare il libero arbitrio.
Naturalmente, i big data offrono numerosi benefici. Ciò che li trasforma in un'arma di disumanizzazione è un limite, non dei big data in sé ma delle modalità con cui li utilizziamo. Il problema è che colpevolizzare le persone per delle azioni che non hanno ancora compiuto significa usare le previsioni ricavate dai big data, e perciò basate sulle correlazioni, per prendere decisioni causali sulla responsabilità individuale.
Forse con un sistema del genere la società sarebbe più sicura o più efficiente, ma una componente essenziale della nostra umanità - la possibilità di scegliere le azioni che intraprendiamo e di risponderne - verrebbe meno. I big data diventerebbero uno strumento per collettivizzare la scelta umana e cancellare il libero arbitrio.
Naturalmente, i big data offrono numerosi benefici. Ciò che li trasforma in un'arma di disumanizzazione è un limite, non dei big data in sé ma delle modalità con cui li utilizziamo. Il problema è che colpevolizzare le persone per delle azioni che non hanno ancora compiuto significa usare le previsioni ricavate dai big data, e perciò basate sulle correlazioni, per prendere decisioni causali sulla responsabilità individuale.
Peccato che gli esseri umani siano naturalmente inclini a vedere il mondo attraverso la lente del rapporto di causa ed effetto. Dunque i big data rischiano costantemente di essere distorti per fini causali, di essere associati a visioni rosee in ordine alla molto maggiore efficacia con cui potremmo assegnare la colpevolezza se potessimo contare sulle previsioni che ne derivano.
Per decenni, uno dei principi ispiratori delle leggi che tutelano la privacy in tutto il mondo è stato lasciare il controllo agli individui, consentendo loro di decidere se, come e da parte di chi potevano essere processate le proprie informazioni personali. Nell'era di Internet, questo ideale encomiabile si è trasformato spesso in un sistema burocratico di «consenso informato». Ma nell'era dei big data, quando il grosso del valore dei dati sta in utilizzi secondari che non si potevano nemmeno immaginare al momento della raccolta, un meccanismo di questo tipo per la difesa della privacy non è più sostenibile.
Noi immaginiamo un quadro di riferimento molto diverso per la tutela della privacy nell'era dei big data, meno focalizzato sul consenso individuale al momento della raccolta e più incentrato sulla responsabilizzazione degli utilizzatori per quello che

fanno. In quel mondo futuribile, le aziende valuteranno formalmente un determinato riutilizzo dei dati in base all'impatto che produce sulle persone di cui si processano informazioni private.

Ciò non dovrà essere meticolosamente dettagliato in tutti casi, perché le future leggi sulla privacy definiranno macro-categorie di utilizzi, inclusi quelli che saranno ammissibili con o senza delle salvaguardie limitate e standardizzate. Per iniziative più rischiose, i regolatori fisseranno dei principi di base su come gli utilizzatori dei dati dovrebbero valutare i pericoli di un uso pianificato e stabilire quale tra questi evita o riduce il danno il più possibile. Una previsione di questo tipo promuove i riutilizzi creativi dei dati, e assicura nel contempo che si prendano misure in grado di escludere danni agli individui[197].

Effettuando correttamente una valutazione dell'utilizzo dei big data e implementandone accuratamente i risultati, si offrono benefici tangibili a chi se ne avvale: in molti casi saranno liberi di perseguire usi secondari dei dati personali, senza dover tornare dai singoli proprietari per ottenerne l'esplicito consenso. Per contro, valutazioni superficiali o un'implementazione inadeguata delle salvaguardie esporranno gli utilizzatori dei dati a una responsabilità giuridica ben precisa, e a provvedimenti dei regolatori come diffide, multe o addirittura procedimenti penali. La responsabilizzazione di chi utilizza i dati funziona solo quando è accompagnata da un sistema sanzionatorio.

Trasferire l'onere della responsabilità dal pubblico agli utilizzatori dei dati appare logico per diverse ragioni. Sanno molto meglio di chiunque altro, e certamente meglio dei consumatori o dei regolatori, come intendono usare i dati. Effettuando direttamente la valutazione (o affidandola a degli esperti), eviteranno il problema di rivelare strategie confidenziali a possibili concorrenti. Ma soprattutto, gli utilizzatori dei dati si accaparrano quasi tutti i benefici dell'uso secondario, perciò è giustissimo chiamarli a risponderne delle loro azioni e pretendere che siano loro a effettuare questa valutazione preventiva su di esse.

Con questo approccio alternativo alla tutela della privacy, gli utilizzatori dei dati non saranno più tenuti per legge a cancellare le informazioni personali quando avranno

[197] BERK R., *op. ult. cit.*, p. 236.

raggiunto il loro scopo primario, come prevedono attualmente quasi tutte le leggi sulla privacy.

Oltre a immaginare l'evoluzione del principio normativo dalla «privacy su consenso» alla «privacy tramite responsabilizzazione», prevediamo che in certi casi l'innovazione tecnica contribuirà a proteggere la privacy. Un approccio che sta emergendo lentamente è il concetto di «privacy differenziale». Si tratta di confondere deliberatamente i dati in modo che la singola query di un grande dataset non riveli i risultati esatti ma solo delle approssimazioni. Ciò rende difficile e costoso associare determinati data point a determinate persone[198].

A prima vista, l'approssimazione deliberata delle informazioni potrebbe far perdere di vista delle indicazioni preziose. Ma non è obbligatorio - o quantomeno -, il compromesso può essere favorevole. Per esempio, gli esperti di politica tecnologica osservano che Facebook si affida a una forma di privacy differenziale quando riporta informazioni sui suoi utilizzatori ai potenziali investitori pubblicitari: i numeri che riporta sono approssimati, per cui non possono aiutare a rivelare le singole identità[199].

L'evoluzione dei meccanismi di controllo dal consenso individuale alla responsabilità di chi utilizza i dati è un cambiamento radicale ed essenziale che appare necessario per una governance efficace delle grandi masse di dati. Ma non è l'unico.

[198] *Ibidem*, p. 238.
[199] *Ibidem*, p. 239.

CAPITOLO TERZO

DATI PERSONALI E BIG DATA: PROFILI NORMATIVI

1. Introduzione

È constatazione oramai diffusa e condivisibile che, nell'attuale fase di evoluzione del sistema economico e sociale, i dati assumono una rilevanza sempre maggiore. Al centro della scena si colloca indubbiamente il fenomeno dei Big Data: vale a dire, le diverse e sempre più sofisticate forme di raccolta, organizzazione ed elaborazione statistico-informatica di enormi quantità di dati (personali o di altro tipo), da parte di imprese e/o pubbliche amministrazioni, al fine di costruire modelli predittivi della dinamica di comportamenti umani e/o o di fenomeni di altro tipo (industriali, naturali, ecc.).

Si tratta di un fenomeno che permette alle imprese e alle amministrazioni pubbliche di migliorare in vario modo la varietà, la qualità e l'efficienza dei prodotti e/o dei servizi offerti alla collettività. Al contempo, i Big Data sollevano delicati interrogativi di ordine politico, etico e giuridico. Tra i numerosi risvolti giuridici del fenomeno, tratteremo qui il tema del regime di "appartenenza" dei dati.

Il tema è delicato e non facile. Come in tutte le situazioni nelle quali si discute del riconoscimento di un diritto di proprietà (e quindi di un diritto di esclusiva) su una risorsa immateriale come le informazioni, anche con riferimento al diritto sui dati, riemerge il dilemma tra l'esigenza di proteggere e remunerare adeguatamente gli investimenti e le attività poste in essere dall'impresa (e/o dalla pubblica amministrazione) per la produzione dei dati contro fenomeni di *free riding*[200], e la contrapposta esigenza dei singoli e/o della collettività di accedere e utilizzare tali dati per le ragioni più diverse (di innovazione e concorrenza, di tutela della *privacy*, di

[200] Cfr., di recente, CIANI J., *Property rights model v. contractual approach: how protecting non personal data in cyberspace?*, in *Dir. Comm. Int.*, 2017, pp. 831 ss.: "*Data-driven innovation can require significant time and up-front investments. These include the costs related to (i) datafication, (ii) data collection, (iii) data cleaning and (iv) data curation. Other than data, a number of complementary resources ranging from data models and algorithms to secured IT infrastructures for data storage, processing, and access may be required. Therefore, from the business perspective, the protection of data may be needed to secure this economic investment*".

diffusione della conoscenza, ecc.). La soluzione di questo nodo non è facile da fornire perché, al momento, non si dispone né a livello europeo né a livello nazionale di regole che disciplinino specificamente il rapporto tra Big Data e proprietà intellettuale, né pare lecito attendersi a breve interventi dal legislatore, essendo estremamente controversa l'opportunità politica dell'introduzione di un diritto di proprietà intellettuale *ad hoc* sui dati[201].

Non è questa la sede per fornire risposte definitive su tali problemi di carattere generale, né tanto meno per esprimere valutazioni di politica legislativa.

A questo proposito, ci si limita a rilevare, a mo' di premessa delle considerazioni che seguono, che un ordinamento evoluto non può prescindere dal riconoscere una qualche forma di tutela all'interesse del soggetto a remunerare adeguatamente gli investimenti e le attività profusi nella produzione di dati, si tratti di un diritto *ad hoc* oppure di strumenti (non pensati specificamente per il fenomeno dei Big Data ma) comunque utilizzabili, eventualmente in modo cumulativo, per fornire una protezione (anche) a questo. Ove così non fosse, il rischio, non tollerabile, sarebbe quello di una produzione insufficiente di dati: perlomeno le imprese, infatti, non possono fare a meno di orientare i propri investimenti e le proprie attività in funzione delle possibilità di ritorno economico degli stessi e, in mancanza di idonei strumenti di tutela contro il *free riding* e di tecniche di adeguata valorizzazione, i costi e i rischi connessi alla produzione dei dati sarebbero difficilmente sostenibili[202]. Spetterà ad altre discipline risolvere i problemi di garanzia dei diritti di accesso, come la disciplina della *privacy*, il diritto antitrust, il diritto dei mercati finanziari o la disciplina della *public sector information*.

Partendo da questo assunto, con le seguenti pagine si offrirà un censimento delle principali questioni giuridiche delle quali l'operatore (impresa o pubblica amministrazione) dev'essere consapevole ed è costretto a fronteggiare ogni qual volta intenda tutelare e/o valorizzare il patrimonio di dati a sua disposizione. In particolare,

[201] È significativo al riguardo l'esito delle consultazioni pubbliche lanciate dalla Commissione europea in merito all'opportunità di introdurre una regolamentazione europea in tema di proprietà dei dati non personali.

[202] Si aggiunga che il riconoscimento di una qualche forma di protezione sui dati può contribuire, in certa misura, anche alla loro condivisione: la tutela contro accessi o utilizzi non autorizzati dei dati incentiva chi le detiene a trasferirli a terzi mediante contratto.

facendo riferimento agli strumenti di protezione attualmente disponibili nell'ordinamento, ci si soffermerà sulle seguenti questioni:

a) questione della "titolarità" dei dati: stabilire, in altri termini, a chi, e a quali condizioni, spetta un diritto di proprietà sui dati;
b) questione della "circolazione" dei dati: individuare, cioè, quali tecniche sono utilizzabili per trasferire o far utilizzare a terzi il diritto di proprietà sui dati.

Questo censimento è il primo *step* della riflessione, necessario per prendere consapevolezza delle implicazioni legali del ricorso ai Big Data. Al contempo, si abbozzeranno delle prime indicazioni operative, utili ad affrontare in modo accorto alcune tra le più delicate delle sfide che si trovano davanti gli operatori che vogliano approfittare di queste nuove opportunità offerte dalle *information and communication technologies*.

2. Un passo indietro: in che senso è possibile parlare di "proprietà" dei dati?

Qualsiasi discorso su titolarità e circolazione della proprietà dei dati non può fare a meno di indicare in che senso è possibile parlare di un diritto di proprietà sui dati, pena l'assoluta indeterminatezza del ragionamento.

In mancanza di una disciplina specifica del fenomeno dei Big Data, è inevitabile dover ricorrere a discipline della proprietà intellettuale di carattere generale, operazione non sempre agevole trattandosi di forme pensate per fenomeni diversi o comunque non del tutto sovrapponibili a quello qui esaminato.

In verità, per alcuni aspetti o momenti del fenomeno complesso che indichiamo con l'espressione Big Data non è difficile indicare gli strumenti di proprietà intellettuale utilizzabili al fine di assicurare una forma di sfruttamento esclusivo. Questo è vero, in particolare, per il *software* e l'*hardware* impiegati per l'elaborazione dei *data set*: sotto questo profilo, è possibile ricorrere agli strumenti del diritto d'autore e/o del diritto di brevetto, senza particolari difficoltà.

Molto più complesso è invece il tema della proprietà dei dati in sé, sia con riferimento al *data set* alla base del procedimento di elaborazione, sia - sebbene in misura minore - per quel che riguarda i dati risultanti da tale procedimento. Sotto questo profilo la mancanza di una disciplina apposita fa sentire tutto il suo peso.

Al riguardo occorre innanzitutto distinguere i dati personali dai dati di tipo non personale (perché afferenti a fenomeni industriali, tecnici o naturali, o perché anonimizzati).

Nel primo caso, non è certo che si possa configurare un vero e proprio diritto di proprietà sui dati, perché la legge[203] riconosce inderogabilmente alla persona cui sono riferibili (il c.d. interessato) dei diritti che limitano o condizionano la libera trasferibilità dei dati. Ciononostante, la legge ammette che un terzo (impresa o ente) possa, a certe condizioni, essere titolare del "trattamento" di tali dati, il che può consentirgli di conseguire anche il potere di sfruttarli economicamente (nell'ambito commerciale, il trattamento consiste nel marketing diretto, nella profilazione, nella cessione o comunicazione a terzi affinché usino i dati per fini commerciali)[204].

Nel caso dei dati non personali, invece, sono senz'altro possibili forme di appropriazione vera e propria dei dati. Due sono gli strumenti che si prestano immediatamente allo scopo: in primo luogo, il diritto di proprietà intellettuale sulle banche dati e, più precisamente, il diritto *sui generis*, previsto e disciplinato dall'art. 102-*bis*, L. n. 633/1941 (o LDA); in secondo luogo, il segreto aziendale (o commerciale) di cui agli artt. 98 e 99, D.Lgs. n. 35/2004 (codice della proprietà industriale o CPI)[205]. Peraltro, queste tecniche possono essere utilizzate anche con riferimento ai data set composti (anche o solo) da dati personali, per rafforzare il controllo che deriva dalla titolarità del trattamento[206].

Va comunque notato che questi strumenti non si prestano facilmente allo scopo, né consentono di dare copertura a tutte le situazioni-tipo che possono verificarsi nella realtà. Basti rilevare che:

- il diritto *sui generis* sulle banche dati presuppone che il *data set* possa essere qualificato come "banca dati" in senso tecnico-giuridico, e cioè come un insieme di informazioni che risponde a determinati requisiti, il primo dei quali

[203] D.Lgs. n. 196/2003; Reg. UE 2016/679.

[204] BANTERLE F., *Brevi cenni sulla titolarità dei dati comportamentali nei Big Data tra privacy e proprietà intellettuale*, in *AIDA*, 2016, pp. 579 ss.

[205] Cfr., F. BANTERLE, *op. ult. cit*., pp. 589 ss.; OTTOLIA A., *Big Data e innovazione computazionale*, Utet, Torino, 2017, pp. 43 ss.; GALLI C. - BOGNI M., *I requisiti per la tutela IP dei Big Data*, in FALCE V. - GHIDINI G. - OLIVIERI G., *Informazione e Big Data tra innovazione e concorrenza*, Giuffrè Editore, Milano, 2018, pp. 96 ss.; M. LIBERTINI, *Le informazioni aziendali riservate (segreti commerciali) come oggetto di diritti di proprietà industriale*, in *Dir. ind*., 2017, pp. 566 ss.

[206] BANTERLE F., *Brevi cenni sulla titolarità dei dati comportamentali nei Big Data tra privacy e proprietà intellettuale*, op. cit., pp. 580 ss.

consiste nel fatto che si tratti di informazioni disposte in modo metodico o sistematico, e quindi ordinato secondo un qualche criterio: il che non avviene nel caso delle raccolte di dati disposti in modo casuale, che quindi non possono beneficiare della protezione del diritto *sui generis*[207];

- il diritto *sui generis* richiede che il *data set* (qualificabile come banca dati) sia il frutto di un investimento rilevante del costitutore nel conseguimento, nella verifica o nella disposizione dei dati: presupposti che non ricorrono nel caso di investimenti (anche importanti ma) non specificamente finalizzati alla costituzione della banca dati; il che preclude la tutela del diritto *sui generis* per i dati la cui produzione o raccolta siano degli aspetti necessari dell'attività di fornitura di prodotti o servizi svolta dal costitutore[208];
- il diritto *sui generis* riserva al titolare di impedire la riproduzione e il reimpiego della banca dati nel suo complesso o di parti sostanziali della stessa, ma non dà alcuna esclusiva sui singoli dati o comunque su parti non sostanziali dell'aggregato[209];
- il segreto aziendale (o commerciale), da un lato, presuppone che il titolare adotti misure adeguate di segretazione delle informazioni, il che non è sempre possibile o agevole[210]; dall'altro, è applicabile solo a informazioni che abbiano

[207] In proposito, v. ad es. OTTOLIA A., *Big Data e innovazione computazionale*, op. cit., p. 75, il quale osserva come tale aspetto della disciplina delle banche dati "*esclude, in particolare, che si possano considerare proteggibili i meri flussi di dati indistintamente rilevati, per esempio, da un social network o dal comportamento di una macchina o da un fenomeno naturale*".

[208] Il che, però, secondo la migliore dottrina, non esclude necessariamente che possano beneficiare della tutela *sui generis* le banche dati che siano costituite nello svolgimento di un'attività più ampia o diversa (come è tipico dei dati raccolti tramite strumenti *IoT*): in tali casi occorrerà, piuttosto, verificare se il costitutore abbia sostenuto degli investimenti esorbitanti rispetto a quelli strettamente necessari per lo svolgimento dell'attività principale, oppure no (cfr., BERTANI M., *Big Data, proprietà intellettuale e mercati finanziari*, in FALCE V. - GHIDINI G. - OLIVIERI G., *Informazione e Big Data*, op. cit., pp. 35 ss.; *Id*., *Banche dati e appropriazione delle informazioni*, in *Eur. e dir. priv*., 2006, pp. 319 ss.; OTTOLIA A., *Big Data*, op. cit., pp. 80 ss.).

[209] BANTERLE F., *Brevi cenni sulla titolarità dei dati comportamentali nei Big Data tra privacy e proprietà intellettuale*, op. cit., pp. 589 ss.

[210] "*Although there might be very different kinds of data and therefore the costs and the difficulty of keeping them secret might vary, most privately produced and held data can be (and are) kept secret by the data holders. On the contrary, trade secret law could apply very hard when taking external datasets. Gathered or obtained data is often publicly available and once the dataset is published, or disclosed for Big Data analytics or in any other way, the protection can no longer be claimed*": così, CIANI J., *Property rights model v. contractual approach: how protecting non personal data in cyberspace?*, op. cit., pp. 831 ss.; cfr., anche, BANTERLE F., *Brevi cenni sulla titolarità dei dati comportamentali nei Big Data tra privacy e proprietà intellettuale*, op. cit., pp. 591.

un valore commerciale, il che è difficilmente asseribile per i singoli dati che compongono il data set[211].

Pertanto, si possono verificare situazioni nelle quali il *data set* non può accedere o non può accedere con sicurezza alla tutela del diritto *sui generis* né a quella del segreto aziendale. Garantire una qualche forma di protezione ai dati in tali situazioni non è affatto un'operazione banale, anche perché, secondo un orientamento ad oggi consolidato in giurisprudenza e dottrina (anche straniere), non è possibile considerare l'informazione o il dato in sé come oggetti di un diritto di proprietà[212]. In questo contesto, le tecniche più prontamente disponibili sono le seguenti:

a) gli accordi attraverso i quali disciplinare l'accesso e l'utilizzo dei data set, ovviamente applicabili solo nei rapporti con le controparti del contratto (come clienti o partners)[213];
b) i rimedi (richieste di inibitoria, rimozione degli effetti e/o di risarcimento del danno) contro eventuali atti di concorrenza sleale (tipicamente per violazione dei principi di correttezza professionale: art. 2598, n. 3, CC), utilizzabili peraltro solo nei confronti di atti di acquisizione o impiego illeciti posti in essere da (o nell'interesse di) imprese concorrenti[214];
c) le misure tecnologiche di protezione, come strumenti di tutela di fatto contro accessi e/o utilizzi non autorizzati dei dati[215].

---

[211] "*Even more difficult it could be to demonstrate that an individual data has commercial value because it is secret. Indeed, data by itself, if not part of a bigger dataset, is often of low value*": così, ancora, CIANI J., *Property rights model v. contractual approach: how protecting non personal data in cyberspace?*, op. cit., pp. 831 ss.

[212] Cfr., ZENO ZENCOVICH V., voce *Cosa*, in *Dig. disc. priv*., *Sez. civ*., III, 438; CIANI J., *Property rights model v. contractual approach: how protecting non personal data in cyberspace?*, op. cit., pp. 831 ss. Non mancano, peraltro, aperture in giurisprudenza sotto il profilo della tutela penale della proprietà dei dati in sé: alcune pronunce straniere hanno riconosciuto che il download o la cancellazione non autorizzati di dati aziendali da parte del dipendente di una società può integrare il reato di furto ed essere come tal sanzionato: cfr., *Cour de cassation*, 20 *may* 2015, *No*. 14-81336; *Oberlandsgericht* Nürnberg, 23 *Januar* 2013. È fortemente dubbio, peraltro, se tali interpretazioni possano essere estese alla tutela civilistica dei dati.

[213] Cfr., Corte di Giustizia dell'Unione Europea, 15/01/2015, C-30/14, *Ryanair v. PR Aviation*, la quale ha statuito che gli accordi che disciplinano l'accesso a *dataset* che non godano delle protezioni specifiche sulle banche dati non sono soggetti all'applicazione dei limiti alle restrizioni contrattuali a carico dell'utilizzatore della banca dati previste dagli artt. 7 e 15 della Dir. 96/9/CE: su questa base, potrebbero legittimamente restringersi anche gli utilizzi di parti non sostanziali del *dataset*; cfr., anche, BANTERLE F., *Brevi cenni sulla titolarità dei dati comportamentali nei Big Data tra privacy e proprietà intellettuale*, op. cit., p. 594.

[214] BANTERLE F., *Brevi cenni sulla titolarità dei dati comportamentali nei Big Data tra privacy e proprietà intellettuale*, op. cit., pp. 594 ss.

[215] *Ibidem*, p. 595.

Alla luce di queste pur sommarie considerazioni, è evidente che più che di un diritto di proprietà sui dati, dovrebbe parlarsi di più diritti di proprietà sui dati e, per alcune tipologie di *data set*, non è forse corretto parlare di proprietà *tout court*.

3. Titolarità dei dati: titolarità individuale e contitolarità

Comprendere a chi spetti la titolarità dei Big Data è tanto importante quanto difficile. Individuare il titolare del *data set* consente di stabilire chi abbia il diritto di sfruttare economicamente i dati, direttamente (all'interno della propria attività economica, ad es. migliorando l'efficienza dei propri processi produttivi o commerciali), o indirettamente (cedendoli a terzi o dandoli in licenza a costoro). Individuare il titolare del *data set*, d'altra parte, non è compito agevole perché il fenomeno dei Big Data è complesso sia da un punto di vista oggettivo, sia da un punto di vista soggettivo: da un lato, i dati presentano natura variegata (personali, o di altro tipo; grezzi o organizzati; a monte o a valle del processo di *analytics*); dall'altro, i fenomeni misurati dai dati, le tecnologie impiegate per la loro produzione, raccolta, archiviazione, organizzazione o analisi, e gli investimenti destinati al loro finanziamento, possono provenire da soggetti differenti.

Per quanto riguarda i dati personali, occorre distinguere tra titolarità dei dati e titolarità del "trattamento" dei dati. La titolarità dei dati spetta alla persona fisica alla quale si riferiscono (il c.d. interessato). La titolarità del trattamento, invece, spetta al soggetto che ha una base giuridica che gli conferisce il potere di determinare le finalità e i mezzi del trattamento; base giuridica è, di norma, il consenso dell'interessato[216]. L'imputazione della titolarità del trattamento non è sempre agevole: il servizio di *cloud computing* affidato in outsourcing è uno dei casi controversi sotto tale profilo, ma ragionevolmente la titolarità del trattamento deve essere ravvisata in capo al cliente della società di *cloud*, cui invece competerà la qualifica di responsabile del trattamento[217].

Non può escludersi che la titolarità del trattamento dei dati personali competa contemporaneamente a più soggetti. In tali casi, occorre distinguere secondo che la

[216] Ibidem, p. 596.

[217] Cfr., BANTERLE F., *op. ult. cit.*, pp. 585 ss., il quale sottolinea come la qualifica del fornitore come mero responsabile del trattamento implichi che il fornitore non sia legittimato a trattare i dati degli utenti del proprio cliente per una propria finalità in difetto dell'autorizzazione di quest'ultimo.

titolarità in capo a più soggetti sia il frutto del caso, oppure dipende da un accordo tra i titolari. Nel primo caso, ciascuno dei titolari del trattamento potrà decidere finalità e mezzi del trattamento in autonomia, per gli ambiti di rispettiva competenza. Nel secondo caso, invece, si determinerà una situazione di contitolarità, per cui le finalità e i mezzi del trattamento dei dati dovranno essere decisi congiuntamente dai contitolari. A questo proposito, la disciplina di settore prevede alcuni obblighi relativi al contenuto dell'accordo di contitolarità, e una disciplina apposita per quanto riguarda la responsabilità dei contitolari nei confronti dell'interessato[218]. Il problema più delicato è, comunque, quello della disciplina delle decisioni relative a fini e mezzi del trattamento: sotto questo profilo, il riferimento normativo alla necessità di decisioni congiunte dei contitolari rischia di complicare l'attività dei contitolari, sicché è opportuno che l'accordo di contitolarità regoli preventivamente il tema (eventualmente autorizzandosi reciprocamente a sfruttare in autonomia i dati in comune, nei confini definiti per contratto).

Per quanto riguarda i dati di altra natura, occorre distinguere tra *data set* "a monte" del processo di *analytics* (dati grezzi o organizzati ma non ancora analizzati), e dati "a valle" del processo. Per altro verso, occorre distinguere tra dati oggetto di diritti di proprietà intellettuale, e dati la cui protezione si basa esclusivamente sulla disciplina di tutela contro la concorrenza sleale, sugli strumenti contrattuali e, eventualmente, sulle misure fisiche o tecnologiche di protezione.

Limitando per ora il discorso ai dati coperti da diritti di proprietà intellettuale, e cominciando dai *data set* non ancora oggetto di analisi, è da ritenere ragionevolmente che la titolarità spetta a chi ha investito nella produzione, raccolta, archiviazione e/o organizzazione del *data set*. Questo sia con riferimento ai dati protetti tramite diritto *sui generis*, sia con riferimento ai dati protetti tramite segreto aziendale. Per i primi, è la legge a stabilire direttamente che il diritto sui generis spetta a colui che ha effettuato l'investimento necessario per la costituzione della banca dati[219]. Per il diritto sulle informazioni aziendali riservate, la stessa conclusione si raggiunge in via

[218] Art. 26, Reg. UE 2016/679.
[219] Art. 102-*bis*, commi 1, lett. a), e 3, LDA.

interpretativa, in applicazione dei principi estrapolabili dalla disciplina del diritto sui generis[220].

In questa prospettiva, non conta se la tecnologia (di raccolta, archiviazione o organizzazione) sia interna all'organizzazione che ha effettuato l'investimento, sia fornita da terzi in *outsourcing* (come avviene tipicamente, ad esempio, per i servizi di *cloud computing*) o comunque sia di proprietà di terzi (come avviene spesso nel campo dell'*IoT*: si pensi ad esempio ai dispositivi attraverso i quali si raccolgono e trasmettono i dati relativi ad un veicolo, ad un macchinario industriale o a quelli rilevati tramite *beacons*). Non conta, inoltre, se i dati si riferiscano a fenomeni interni all'organizzazione che ha effettuato l'investimento, oppure si riferiscano a terzi (cioè a loro comportamenti, situazioni, posizioni, ecc.: si pensi, ad esempio, ai consumi di energia rilevati tramite *smart meters*). Nei contratti in questione è comunque opportuno chiarire tramite apposite clausole questi aspetti.

Per quanto riguarda la titolarità dei dati o informazioni "a valle" del processo di *analytics*, invece, è ragionevole ritenere che essa spetti a chi ha investito nella elaborazione del *data set*[221]. Anche in questo contesto, il principio è da ritenersi applicabile sia ai dati o informazioni protetti dal diritto *sui generis* sulla banca dati sia quelli coperti dal segreto aziendale. Ancora una volta, non conta se la tecnologia e l'*expertise* impiegate per l'elaborazione del *data set*, sia interna all'organizzazione che ha effettuato l'investimento per il processo di analisi oppure sia fornita da terzi in outsourcing: anche in quest'ultimo caso, la titolarità dei dati o delle informazioni risultati dalla elaborazione dovrà imputarsi a chi ha investito nel processo (salvo pattuizione contraria). Nei contratti in questione è comunque opportuno chiarire tramite apposite clausole questi aspetti.

Anche nel campo dei dati di carattere non personale sono possibili situazioni di contitolarità del *data set*: si pensi, ad esempio, all'ipotesi in cui gli investimenti necessari per la produzione, raccolta, archiviazione o organizzazione del *data set* provengano da imprese differenti; oppure al caso, in cui più soggetti abbiano concorso agli investimenti necessari per l'effettuazione del processo di *analytics* sul *data set*.

---

[220] BANTERLE F., *Brevi cenni sulla titolarità dei dati comportamentali nei Big Data tra privacy e proprietà intellettuale*, op. cit., pp. 593 ss.; BERTANI M., *Proprietà intellettuale e nuove tecniche di appropriazione delle informazioni*, in *AIDA*, 2005, p. 322; OTTOLIA A., *Big Data e innovazione computazionale*, op. cit., pp. 57 ss.

[221] OTTOLIA A., *op. ult. cit*., pp. 252 ss.

In tutte queste ipotesi, si determina una situazione di contitolarità (originaria) sui dati, che è fonte di problemi delicati e complessi: a cominciare dal problema di chi e come può sfruttare economicamente i dati e le informazioni comuni. A differenza di quanto avviene in altre esperienze giuridiche, nell'ordinamento italiano non è prevista una disciplina specifica per le situazioni di contitolarità di diritti di proprietà intellettuale, a parte un rinvio[222] generico e di difficile interpretazione alla disciplina generale della comunione[223], che però è pensata per la proprietà di beni materiali, ed è difficilmente adattabile al contesto di beni immateriali. È così, ad esempio, che, mentre è chiaro che, in una situazione di comunione di un immobile (un terreno, un'abitazione o altro), ciascuno dei comproprietari abbia il diritto di utilizzare liberamente il bene comune (ad es. per passeggiarvi, abitarvi, ecc.) purché non impedisca agli altri di fare altrettanto[224], nel caso dei beni immateriali (come un brevetto, un marchio o, per quel che qui interessa, un *data set*), non è altrettanto semplice stabilire se (ed è probabilmente da escludere che) ciascuna delle imprese contitolari abbia analoga facoltà di impiegare liberamente la risorsa comune all'interno della propria attività oppure debba concordare preventivamente con le altre se, come e in che misura tale sfruttamento possa avvenire. In questi contesti lo sfruttamento della risorsa comune è questione assai più delicata, ad esempio perché una delle imprese ha un'organizzazione e una posizione di mercato tali da permetterle saturare il mercato nel quale può essere sfruttato il bene comune, e lasciando così l'altra impresa contitolare… a bocca asciutta. Anche da questo punto di vista, quindi, è di fondamentale importanza che i contitolari raggiungano un accordo sin da subito al fine di scongiurare possibili conflitti successivi: ad esempio concedendosi delle licenze reciproche che autorizzino e regolamentino lo sfruttamento indipendente dei dati comuni.

Un discorso a parte va fatto per quanto riguarda i dati che non godano della protezione del diritto *sui generis* né di quella del segreto aziendale, quali sono tipicamente i dati "grezzi" non ancora sottoposti a trattamento analitico né organizzati o segretati. In tal caso, non essendo in questione il riconoscimento di un vero e proprio diritto di proprietà intellettuale, il tema della titolarità dei dati diventa di soluzione più

[222] Art. 6 c.p.i.
[223] Artt. 1100 ss. c.c.
[224] Art. 1102 c.c.

complessa. A tal fine, dovrà inevitabilmente farsi riferimento ai caratteri e ai principi tipici dello strumento (giuridico e/o tecnico) cui, di volta in volta, si affida la protezione dei dati. Così ragionando, può allora schematicamente ipotizzarsi che:

- i dati protetti tramite accordi contrattuali che ne disciplinano l'accesso e l'utilizzo spetteranno a chi è individuato come titolare all'interno dell'accordo, e sempre che, in relazione al profilo della titolarità dei dati, l'accordo possa ritenersi espressione di interessi meritevoli di tutela secondo l'ordinamento giuridico. Tale giudizio di meritevolezza sarà questione da risolversi di volta in volta alla luce di una complessiva valutazione delle circostanze del caso concreto. È ragionevole presumere, tuttavia, che debbano ritenersi valide clausole contrattuali che attribuiscano la titolarità dei dati grezzi al soggetto (o ai soggetti) che ha (o abbiano) effettuato gli investimenti necessari per la loro produzione, raccolta e/o archiviazione: alla luce delle considerazioni esposte a proposito dei *data set* protetti da diritti di proprietà intellettuale, può infatti prospettarsi la tesi che l'investimento nelle attività citate esprima un interesse sufficientemente meritevole di tutela secondo l'ordinamento giuridico;
- i dati protetti mediante i rimedi contro gli atti di concorrenza sleale spettino all'impresa nel cui complesso aziendale i dati siano inseriti e che abbia sostenuto i costi necessari alla produzione, raccolta, archiviazione e/o analisi. Anche in questo contesto, pertanto, il profilo dell'investimento nelle attività di Big Data assume probabilmente un ruolo chiave;
- i dati protetti mediante misure tecnologiche di protezione - misure che, in sé e per sé, hanno carattere fattuale e non giuridico - dovrebbero spettare a chi, in concreto, abbia adottato tali misure. È ragionevole ipotizzare, tuttavia, che la titolarità assicurata mediate tali misure non possa condurre a risultati in contraddizione con quelli derivanti da eventuali tutele contrattuali o concorrenziali: è chiaro, ad esempio, che dati acquisiti con modalità illegittime (ad esempio tramite spionaggio industriale, o in violazione di accordi contrattuali pregressi) non potranno considerarsi di titolarità dell'autore dell'illecito o dell'inadempimento sol perché costui abbia avuto l'accortezza di proteggersi mediante misure tecnologiche.

4. Circolazione dei dati: operazioni di scambio e operazioni di cooperazione

I Big Data possono essere sfruttati economicamente all'interno dell'attività del titolare (o dei contitolari), ma anche all'esterno. Innanzitutto questo sfruttamento esterno può avvenire da parte di terze parti cui il titolare affidi lo sfruttamento diretto a fronte di un corrispettivo, nell'ambito di operazioni di scambio[225].

In questo contesto, la circolazione dei dati può avvenire innanzitutto a titolo definitivo: con rinuncia, cioè, del titolare originario a ogni possibilità di sfruttamento diretto o ulteriore trasferimento a terzi; tipicamente a fronte di un corrispettivo *una tantum*. Lo strumento utilizzabile a tal fine è il contratto di cessione che dovrà avere ad oggetto il diritto o i diritti di proprietà intellettuale sul *data set*. Occorre precisare, peraltro, che nel caso dei dati personali, una vera e propria cessione sarà possibile solo con riferimento alla titolarità del trattamento dei dati, mentre nessuna forma di rinuncia definitiva del titolare è ammissibile per quel che riguarda i dati in sé e per sé da parte dell'interessato[226].

La circolazione può anche avvenire a titolo provvisorio: attribuendo alla controparte solo il diritto di sfruttare economicamente il *data set*, a fronte di un compenso periodico commisurato ai guadagni conseguiti dal *partner* nello sfruttamento della risorsa, e conservando al titolare il diritto di sfruttare e/o di disporre ulteriormente della risorsa concessa. Lo strumento contrattuale utilizzabile per regolare tali operazioni sarà tipicamente un contratto di licenza, che potrà prevedere o meno un diritto di esclusiva a favore del licenziatario: secondo come viene confezionato, il licenziatario potrà pretendere che il titolare non dia in licenza o non ceda a terzi il diritto di sfruttare il *data set*, o anche che si astenga egli stesso dallo sfruttare direttamente la risorsa nella propria attività economica. Simili forme di licenza esclusiva sono senz'altro ammissibili con riferimento ai dati di carattere non personale, e devono probabilmente ammettersi anche con riguardo ai dati personali[227]. Non è da escludere, del resto, la possibilità di ricorrere a licenze non esclusive a titolo gratuito, sul modello delle licenze *creative commons*: in questo modo, può formalizzarsi un regime di *open data*, sulla falsariga di quanto avviene da tempo nel campo del *software* e delle opere letterarie.

---

[225] OTTOLIA A., *Big Data e innovazione computazionale*, op. cit., pp. 221 ss.
[226] *Ibidem*, pp. 227 ss.
[227] *Ibidem*, pp. 230 ss.

Sono poi possibili anche forme di collaborazione[228]. Simili forme di cooperazione sono particolarmente appetibili nel contesto delle piccole e medie imprese, le quali possono aver interesse a condividere i rispettivi *data set* e/o le rispettive risorse finanziarie, organizzative o tecnologiche, al fine di rendere possibili forme di *analytics* che, da sole, non sarebbero in grado di realizzare o che possono realizzare in modo più efficiente in forma collaborativa. Da un punto di vista giuridico, simili operazioni possono essere strutturate in forma societaria o consortile: in tali casi sarà fondamentale confezionare gli atti costitutivi dell'organizzazione in modo da garantire un equilibrio nella *governance* e nella fruizione delle risorse e dei servizi messi in comune. Non può escludersi nemmeno che simili forme di cooperazione diano luogo a situazioni di formale contitolarità (derivativa) dei *data set*: in tal caso, si riproporranno le spinose questioni già evidenziate a proposito della contitolarità originaria, e sarà ancora più importante dedicare la massima attenzione e cura alla redazione del testo del contratto, onde prevenire possibili conflitti che potrebbero essere fatali per la programmata collaborazione.

È il caso di segnalare, inoltre, che sia nel campo delle operazioni di scambio, sia nel campo delle operazioni di collaborazione, la circolazione dei *data set* può avvenire non solo in forma isolata, ma anche nel contesto di più ampie operazioni di trasferimento d'azienda o di ramo d'azienda (a titolo definitivo o provvisorio). In tali casi, la cessione o la licenza sul *data set* costituirà un elemento del contratto di cessione, affitto o conferimento del complesso aziendale: e sarà opportuno esplicitare nel testo del contratto l'estensione dell'operazione al *data set*, al fine di prevenire possibili controversie tra le parti sul punto, alimentate dalla difficoltà di qualificare i dati come "beni" aziendali, soprattutto nel caso in cui i dati non siano protetti mediante diritti di proprietà intellettuale. Naturalmente, in tali casi dovranno applicarsi le regole specifiche per il trasferimento d'azienda (iscrizione del contratto nel registro delle imprese, divieto di concorrenza per l'alienante, responsabilità dell'acquirente per i debiti derivanti dalla gestione pregressa, ecc., secondo quanto previsto dagli artt. 2555 ss. c.c.).

Da ultimo, è importante sottolineare che, quale che sia il tipo di operazione circolatoria programmata (di scambio o di collaborazione; isolata o inclusa nel trasferimento di un

[228] *Ibidem*, pp. 239 ss.

complesso aziendale), sarà fondamentale che, nel formalizzare contrattualmente la stessa, vengano adottati gli opportuni accorgimenti per preservare la situazione proprietaria oggetto di negoziazione e la funzione dell'operazione stessa. In particolare, il problema si pone per i dati protetti tramite segreto: in tal caso, sarà fondamentale assicurarsi, tramite apposite convenzioni di riservatezza, che il cedente o, secondo i casi, il licenziatario, si impegnino a non divulgare, consentire l'accesso o l'utilizzo a terzi non autorizzati dal titolare, pena la perdita del diritto di proprietà intellettuale[229]. Il problema si pone, altresì, per le operazioni di scambio aventi ad oggetto *data set* che non godano della protezione del diritto *sui generis* né di quella del segreto aziendale: in questi casi, occorrerà inserire nel contratto di cessione il divieto per l'alienante di sfruttare i dati al fine di preservare il senso economico-giuridico della operazione.

Per altro verso, sarà opportuno adottare specifiche previsioni contrattuali con riferimento al rischio che i dati oggetto di trasferimento non siano corretti e/o non corrispondano a quando convenuto tra le parti. A tal fine, si potrebbe ricorrere a formule analoghe a quelle tipicamente utilizzate nelle *declarations and warranties* nelle operazioni di trasferimento d'azienda o di partecipazioni societarie.

Inoltre, le operazioni circolatorie aventi ad oggetto i dati dovranno rispettare le regole imperative di volta in volta eventualmente applicabili. Con particolare frequenza il problema riguarderà la disciplina in tema di *privacy*, la quale pone dei vincoli alla circolazione dei dati personali per effetto dei diritti riconosciuti all'interessato: si pensi, ad esempio, al diritto alla portabilità dei dati o, ancora, al diritto alla revoca del consenso al trattamento. Non può escludersi nemmeno l'eventualità che trovino applicazione regole a tutela della parte contraente "debole", anche nei rapporti contrattuali tra imprese: il problema può presentarsi in situazioni in cui un'impresa sia in condizione di dipendenza economica e tecnologica dalla controparte contrattuale, come non è raro accada nei rapporti tra cliente e provider del servizio di *cloud computing* e/o di *analytics*. In tali casi, troverà applicazione il divieto di abuso di dipendenza economica[230] e, pertanto, occorrerà fare attenzione a che il contratto di

[229] Sottolinea l'importanza delle clausole di riservatezza, con particolare riferimento ai contratti con il *cloud provider*, BANTERLE F., *Brevi cenni sulla titolarità dei dati comportamentali nei Big Data tra privacy e proprietà intellettuale*, op. cit., p. 591.
[230] Art. 9, L. n. 192/1998.

fornitura non contenga clausole eccessivamente onerose a carico della parte "debole" (tipicamente il cliente), anche sotto il profilo della attribuzione della titolarità dei dati[231].

5. Dati personali e Big Data: la "nuova" *privacy* europea

«La rapidità dell'evoluzione tecnologica e la globalizzazione comportano nuove sfide per la protezione dei dati personali. La portata della condivisione e della raccolta di dati personali è aumentata in modo significativo. La tecnologia attuale consente tanto alle imprese private quanto alle autorità pubbliche di utilizzare dati personali, come mai in precedenza, nello svolgimento delle loro attività. Sempre più spesso, le persone fisiche rendono disponibili al pubblico su scala mondiale informazioni personali che le riguardano. La tecnologia ha trasformato l'economia e le relazioni sociali e dovrebbe facilitare ancora di più la libera circolazione dei dati personali all'interno dell'Unione e il loro trasferimento verso paesi terzi e organizzazioni internazionali, garantendo al tempo stesso un elevato livello di protezione dei dati personali»[232].

In precedenza si è dato atto di come il fenomeno della digitalizzazione abbia portato alla creazione di un nuovo ecosistema, nel quale il concetto di "dato" assume un ruolo centrale, costituendo, per così dire, la forma di vita tipica e primordiale.

E se è apparso da subito complicato attribuire una definizione e decifrare con esattezza tutte le caratteristiche che accumunano il *genus* dei "dati", è quasi paradossale pensare che una *species* al suo interno abbia da anni dei contorni molto definiti e sia stata oggetto di specifica ed ampia normazione a livello sia comunitario che nazionale. Il riferimento va ovviamente a quella categoria di dati che risponde al nome di "dati personali", sul cui ambito e sulla cui definizione si tornerà approfonditamente, valutando se e in che modo la disciplina giuridica applicabile, meglio nota come *privacy*, possa avere un impatto sul tema dei Big Data o possa essere da questi influenzata e interessata.

[231] OTTOLIA A., *Big Data e innovazione computazionale*, op. cit., pp. 247 ss.
[232] Considerando 6, Reg. UE 2016/679.

Quanto alle fonti che si andranno a esaminare, il punto di partenza è rappresentato senza dubbio dal nuovo Regolamento Generale sulla Protezione dei Dati[233], già entrato in vigore e applicabile in tutti gli Stati membri dal 25/05/2018.

Non si potrà ovviamente prescindere dalla normativa nazionale ancora vigente (su tutti il D.Lgs. 30/06/2003, n. 196, meglio noto e di seguito indicato come "Codice *Privacy*"[234]) e, per una interpretazione autentica dei principi in materia, dalle opinioni e raccomandazioni del Garante Europeo della Protezione dei Dati (GEPD) e del Gruppo di Lavoro ex art. 29 della Dir. 95/46/CE, organismo consultivo e indipendente, composto da un rappresentante delle autorità di protezione dei dati personali designate da ciascuno Stato membro, dal GEPD, nonché da un rappresentante della Commissione.

Resta inteso che l'obiettivo di questa pubblicazione non è certo quello di esaurire ogni aspetto legato alla protezione dei dati personali né affrontare nel dettaglio tematiche che richiederebbero approfondimenti ben maggiori e incompatibili con lo scopo del presente lavoro. Ciò che si cercherà di mettere in luce sarà piuttosto l'impatto che il fenomeno della digitalizzazione e dei Big Data ha avuto sull'approccio alla *privacy*, così come già in parte recepito dal legislatore europeo e dai Garanti degli stati membri.

---

[233] Reg. UE 2016/679 del Parlamento Europeo e del Consiglio del 27/04/2016 relativo alla protezione delle persone fisiche con riguardo al trattamento dei dati personali, nonché alla libera circolazione di tali dati e che abroga la Dir. 95/46/CE, di seguito anche "GDPR" o "Regolamento"

[234] Sul punto si segnala che, con la Legge di delegazione europea 2016-2017 del 25/10/2017, n. 163 (art. 13), il Governo è stato incaricato di adeguare la normativa nazionale al GDPR e, nello specifico, abrogare espressamente le disposizioni del Codice Privacy incompatibili con il Regolamento, ovvero modificare il Codice Privacy limitatamente a quanto necessario per dare attuazione alle disposizioni non direttamente applicabili, coordinando le disposizioni vigenti e adeguando l'impianto sanzionatorio penale e amministrativo. La prima proposta di Decreto Legislativo di armonizzazione, emanata dal Governo in data 21/03/2018 ha raccolto diverse critiche sia in dottrina che tra le principali associazioni di categoria coinvolte, rallentando inevitabilmente l'iter legislativo. Nel mese di maggio 2018, la seconda proposta ha quindi abbandonato l'iniziale intenzione di abrogare indiscriminatamente la previgente legislazione, per optare verso una sua – più coerente – revisione sistematica in ragione delle novità introdotte dal GDPR ed una modifica del Codice Privacy (da mantenersi quindi nella sua emendata versione). Il termine inizialmente previsto al 21/05/2018 per l'adozione del testo definitivo del decreto di armonizzazione è stato, da ultimo, posticipato al 21/08/2018. Alla data di completamento del presente lavoro, 25/05/2018, si riportano quindi le indicazioni ricavabili dall'ultima proposta circolata, rimandando, per successivi aggiornamenti sulla base degli interventi normativi che verranno effettuati in materia, alla sezione dedicata sul sito *http://info.wolterskluwer.it/gdpr*.

6. La definizione di "dato personale": i *Big Personal Data*

Capire la natura dei dati trattati è una questione dirimente e preliminare a qualsiasi operazione, essendo molteplici le implicazioni che da tale valutazione possono discendere.

Come detto, se manca una disciplina organica sul "dato", non altrettanto si può dire del dato "personale", oggetto della ben nota branca del diritto che va sotto il nome comune di *privacy*. Ma cosa si intende per "dato personale"? È davvero univoca tale definizione? Si cercherà di dare una rapida ma, per quanto possibile, esaustiva disamina della nozione normativamente prevista ed oggi codificata all'art. 4(1) del GDPR in base al quale per dato personale si intende "qualsiasi informazione riguardante una persona fisica identificata o identificabile («interessato»)". Il tutto nella prospettiva di vedere poi quali sono le peculiarità applicative poste da quelli che potremmo chiamare "*big personal data*".

In questo compito si utilizzeranno come riferimenti le linee guida del Parere 4/2007 sul concetto di dati personali adottato il 20/06/2007 dal Gruppo di Lavoro ex art. 29, a cui si faranno necessari rimandi, ove opportuno.

6.1 "Qualsiasi informazione …"

Il primo elemento della definizione sopra riportata è di certo il più volutamente ampio, concretizzando una precisa intenzione del legislatore di ricomprendere nella nozione in commento ogni tipo di affermazione su una persona. Sono così ricomprese informazioni "oggettive" legate a dati di fatto (per es. l'età anagrafica, l'indirizzo, il numero di telefono mobile, la presenza di un certo quantitativo di sostanze nel sangue di un paziente), ma anche "soggettive", quali valutazioni, opinioni o pareri riconducibili ad un dato individuo.

Dal punto di vista del contenuto, un dato personale può ricomprendere informazioni riguardanti la sfera privata, quella familiare o lavorativa, i rapporti economici o sociali di una persona ed ogni possibile interazione con l'esterno, indipendentemente dalla qualifica ricoperta dal soggetto interessato.

Il tutto a prescindere che si tratti di affermazioni vere o inventate ed indifferentemente rispetto al formato (grafico, alfanumerico, fotografico, acustico) o al supporto (cartaceo, digitale, audiovisivo, …) mediante cui queste affermazioni sono rese

disponibili. Proprio sotto questo profilo il concetto di informazione trascende e preesiste alla distinzione tra dato personale "digitale" e "analogico".

6.2 "… riguardante …"

Se il primo concetto è risultato *prima facie* omnicomprensivo e di immediata intuizione, per poter parlare di dati personali occorre in aggiunta che le informazioni "riguardino" (ossia in altre parole siano pertinenti) un determinato soggetto.

Vi saranno molti casi in cui l'applicazione di questo requisito non porrà problemi di sorta (si pensi al nome di un cliente, all'età di un paziente, all'immagine di una persona registrata), ma ciò non esclude un'ampia casistica dal carattere più controverso (e la possibilità di trattare Big Data che contengono miriadi di informazioni non farà altro che aumentare il panorama dei possibili esempi). Si considerino i dati trasmessi o riferiti in primo luogo ad oggetti, quale il valore di un immobile che in sé costituisce un'informazione su un bene e pertanto prescinde dall'applicazione diretta della normativa in tema di protezione dei dati personali, ma che in talune circostanze può rappresentare un dato su un determinato soggetto ad esempio per determinare la tassazione del proprietario dell'immobile stesso.

Per poter stabilire se le informazioni "riguardano" un determinato soggetto occorre che sia fatto riferimento, alternativamente, ad uno dei seguenti elementi:

- contenuto, ossia l'informazione riguarda un determinato soggetto a prescindere dalla finalità del titolare del trattamento o di terzi o dall'impatto sulla persona interessata;
- finalità, ossia l'informazione è utilizzata dal titolare al fine di valutare, trattare in uno specifico modo o influire sullo stato o sul comportamento di un soggetto interessato;
- risultato, ossia l'informazione, pur in assenza di elementi di contenuto o finalità, concerne una persona in quanto il suo impiego può avere un impatto sui diritti ed interessi di quel soggetto.

6.3 "… una persona fisica …"

La normativa in tema di *privacy* compie una scelta di campo piuttosto netta, giungendo a circoscrivere i dati personali a tutte quelle informazioni che riguardano una "persona fisica", a prescindere dalla sua nazionalità o luogo di residenza. Ciò in applicazione di più ampi principi di diritto che trovano la loro fonte nella Dichiarazione universale dei diritti dell'uomo, il cui art. 6 testualmente afferma che "ogni individuo ha diritto, in ogni luogo, al riconoscimento della sua personalità giuridica".

La presa di posizione del GDPR[235] ignora quell'apertura che la precedente normativa di rango europeo aveva avallato a favore delle persone giuridiche e che aveva visto alcune legislazioni nazionali (quali quella italiana poi riformata con l'art. 40, comma 2, lett. b, D.L. 06/12/2011, n. 201, convertito, con modificazioni, dalla L. 22/12/2011, n. 214) ricomprendere nella nozione di "interessato" non solo la persona fisica cui si riferiscono i dati personali, ma anche "la persona giuridica, l'ente o l'associazione" ed estendere a tali soggetti alcune disposizioni in materia di comunicazioni elettroniche[236].

6.4 "… identificata o identificabile, direttamente o indirettamente …"

Il requisito in parola è quello che più risente dello sviluppo tecnologico ed è maggiormente interessato dal dibattito sulla digitalizzazione e sull'avvento dei Big Data, ossia sulla possibilità da parte degli operatori di avere a disposizione sempre più informazioni che, tra loro messe in relazione, assumono un valore maggiormente degno di attenzione.

Un soggetto può dirsi "identificato" se è distinto all'interno di un gruppo da tutti gli altri membri. Parimenti è "identificabile" quando questa distinzione, ancorché non ancora avvenuta, sia comunque possibile.

---

[235] Il cui Considerando 14 ora esplicita che "Il presente regolamento non disciplina il trattamento dei dati personali relativi a persone giuridiche, in particolare imprese dotate di personalità giuridica, compresi il nome e la forma della persona giuridica e i suoi dati di contatto".

[236] Per completezza si segnala come l'attuale normativa europea non si applichi alle persone decedute, pur lasciando aperta la possibilità per gli Stati membri di prevedere una normativa ad hoc sul punto (Considerando 27), mentre non è possibile cristallizzare un approccio uniforme in tema di nascituri. Proprio in tema di persone decedute, la proposta di Decreto Legislativo di armonizzazione del maggio 2018 ha proposto l'introduzione nel Codice Privacy dell'art. 2-duodecies, prevedendo che i diritti dell'interessato, riferiti a dati personali concernenti persone decedute, possano essere esercitati da coloro che abbiano un interesse proprio o agiscano a tutela dell'interessato, in qualità di suoi mandatari, o "per ragioni familiari meritevoli di protezione", salva un'eventuale volontà contraria dell'interessato all'esercizio di tali diritti che sia specifica, libera, informata e risultante in modo non equivoco.

Per quanto riguarda il concetto di "direttamente", il riferimento va ovviamente e primariamente al nome di un dato individuo, pur potendo essere necessario, a seconda del contesto, combinare altre informazioni quali data di nascita, indirizzo, fotografia del volto, ed evitare così omonimie. Con riferimento, invece, alle persone identificate o identificabili "indirettamente", questa categoria rimanda tipicamente al fenomeno delle "combinazioni uniche", siano esse ampie o ridotte. Laddove gli identificatori disponibili non consentono in maniera immediata di identificare un dato soggetto, questi potrà comunque essere considerato "identificabile" laddove le informazioni combinate con altre (che siano o meno conservate dal titolare) consentiranno di distinguerlo dagli altri. Le persone fisiche possono essere associate a identificativi *online* generati dai dispositivi, dalle applicazioni, dagli strumenti e dai protocolli utilizzati, quali gli indirizzi IP, a marcatori temporanei (cookies) o a identificativi di altro tipo; tali identificativi possono lasciare tracce che, se combinate con identificativi univoci e altre informazioni ricevute dai server, possono essere utilizzate per creare profili delle persone fisiche e identificarle[237].

Per stabilire l'identificabilità di una persona è opportuno considerare tutti i mezzi e i fattori obiettivi di cui il titolare del trattamento o un terzo può ragionevolmente avvalersi: tra questi i costi e il tempo necessario per l'identificazione, tenendo conto sia delle tecnologie disponibili al momento del trattamento, sia degli sviluppi tecnologici. È evidente che tale concetto, ben riassunto nel Considerando 26 del GDPR[238], andrà sempre più calato nel contesto tecnologico di riferimento e non si può non ipotizzare che l'utilizzo e la disponibilità di Big Data, come anticipato, portino ad impattare in maniera significativa sulla nozione di "identificabilità", posto il maggior accesso alle informazioni personali da parte dei player digitali a costi contenuti e in tempi sempre più ridotti.

Si consideri, ad esempio, che gli indirizzi IP sono stati ritenuti dati concernenti una persona identificabile, sul presupposto che "i fornitori di accesso Internet e i gestori

[237] Considerando 30.

[238] In base al quale "Per stabilire l'identificabilità di una persona è opportuno considerare tutti i mezzi, come l'individuazione, di cui il titolare del trattamento o un terzo può ragionevolmente avvalersi per identificare detta persona fisica direttamente o indirettamente. Per accertare la ragionevole probabilità di utilizzo dei mezzi per identificare la persona fisica, si dovrebbe prendere in considerazione l'insieme dei fattori obiettivi, tra cui i costi e il tempo necessario per l'identificazione, tenendo conto sia delle tecnologie disponibili al momento del trattamento, sia degli sviluppi tecnologici".

delle reti LAN possono, utilizzando mezzi ragionevoli, identificare gli utenti Internet cui essi hanno attribuito indirizzi IP, poiché, normalmente, essi "registrano" in un apposito *file* la data, l'ora, la durata e l'indirizzo IP dinamico assegnato all'utente Internet. Lo stesso dicasi per i fornitori di servizi Internet, i quali detengono un registro sul *server* HTTP. In questi casi, non vi è dubbio sul fatto che si possa parlare di dati personali"[239].

Sul tema in oggetto, giova altresì menzionare quanto riportato dal Garante Europeo della Protezione dei Dati, che, analizzando le "sfide" poste dai Big Data, ha correttamente osservato quanto segue, con specifico riferimento all'economia digitale e alle società operanti sul web e alla loro possibilità di "identificare" singoli soggetti: «*One result is the emergence of a revenue model for Internet companies relying on tracking online activity. Such 'Big Data' should be considered personal even where anonymisation techniques have been applied: it is becoming and will be ever easier to infer a person's identity by combining allegedly 'anonymous' data with publicly available information such as on social media. Furthermore, with the advent of the 'Internet of Things', much of the data collected and communicated by the increasing number of personal and other devices and sensors will be personal data: the data collected by them can be easily related to the users of these devices whose behaviour they will monitor. These may include highly sensitive data including health information and information relating to our thinking patterns and psychological make-up*»[240].

7. Operare correttamente con i *Big Personal Data*

7.1 I presupposti e il fondamento giuridico del trattamento

Chiarito cosa si intenda per "dato personale", la valutazione necessariamente si sposta sui presupposti e sulle condizioni di liceità del trattamento. Lavorare con o attraverso *big personal data* per un operatore, infatti, a prescindere dal numero e dall'entità dei dati in questione, significa *in primis* conoscere i fondamenti che giustificano un

[239] Cfr. Documento di lavoro: *Tutela della vita privata su Internet - Un approccio integrato dell'EU alla protezione dei dati on-line adottato dal Gruppo di Lavoro ex art. 29 il 21/11/2000* (5063/00/IT/DEF - WP 37).

[240] *Opinion 7/2015* adottata dal Garante europeo della protezione dei dati il 19/11/2015 - *Meeting the challenges of Big Data - A call for transparency, user control, data protection by design and accountability*.

determinato trattamento e comprendere entro quali limiti considerare lecita (o illecita) una specifica attività di raccolta, elaborazione, archiviazione e circolazione di dati.
Si ricorda che i dati trattati in violazione della disciplina in materia sono "inutilizzabili" (ciò verrebbe, peraltro, formalmente esplicitato dal nuovo art. 2-novies del Codice *Privacy*, come da proposta di Decreto Legislativo di armonizzazione del maggio 2018) e possono portare all'applicazione di sanzioni di natura amministrativa o penale a seconda della fattispecie.
A questo proposito e con specifico riferimento al tema dei Big Data, si sottolinea la recente proposta di introduzione nel Codice *Privacy*, ad opera del Decreto Legislativo di armonizzazione del maggio 2018 di cui sopra, di due distinte fattispecie di reato aventi ad oggetto dati personali "riferibili a un rilevante numero di persone"[241], di cui vengono sanzionate:

- la comunicazione e diffusione illecita (Art. 167-*bis*), punibili con la reclusione da uno a sei anni;
- l'acquisizione fraudolenta (Art. 167-*ter*), punibile con la reclusione da uno a quattro anni.

7.2 Il trattamento secondario compatibile: il concetto di *purpose limitation*

Risulta evidente come le tecnologie che si basano sull'elaborazione ed analisi dei Big Data moltiplichino le potenzialità dei dati e i loro utilizzi, aumentando esponenzialmente le possibili attività che un operatore può mettere in campo partendo da una base comune di informazioni, più o meno vaste dal punto di vista numerico e qualitativo. Questa considerazione, se da un punto di vista tecnologico rappresenta uno dei motivi di maggior interesse ed approfondimento nella costante ricerca di una evoluzione che soddisfi quante più esigenze possibili del mercato digitale, sotto il profilo della *privacy* finisce per scontrarsi con uno dei principi cardine della normativa applicabile, che ha trovato anche da ultimo una sua regolamentazione a livello comunitario nel GDPR, ossia il principio di limitazione delle finalità. Tale principio

[241] Come precisato nella Relazione Illustrativa alla proposta di Decreto Legislativo di armonizzazione del maggio 2018, "l'aggettivo rilevante, già adoperato dal legislatore penale, è stato riferito al numero delle persone offese, per modo che indirettamente vengono sanzionate condotte che riguardano una quantità considerevole di dati personali". Ciò al fine di reprimere quei comportamenti che, per vastità di dimensioni, coinvolgimento di un numero ingente di soggetti (e quindi di dati trattati), finalizzazione dell'azione al profitto, non si esauriscono nella mera violazione delle norme sul trattamento.

ha trovato una sua codificazione sin dall'art. 6(1)(b) della Dir. 95/46/CE, poi recepito dall'art. 11 del Codice *Privacy* in base al quale "I dati personali oggetto di trattamento sono […] raccolti e registrati per scopi determinati, espliciti e legittimi, ed utilizzati in altre operazioni del trattamento in termini compatibili con tali scopi".

Analoga formulazione è ora contenuta nell'art. 5(1)(b) del GDPR, secondo cui "I dati personali sono [il testo inglese utilizza l'espressione '*shall be*'] […] raccolti per finalità determinate, esplicite e legittime, e successivamente trattati in modo che non sia incompatibile con tali finalità; un ulteriore trattamento dei dati personali a fini di archiviazione nel pubblico interesse, di ricerca scientifica o storica o a fini statistici non è, conformemente all'art. 89, par. 1, considerato incompatibile con le finalità iniziali («limitazione della finalità»)".

Ora, il principio in parola contiene a ben vedere due diverse indicazioni che meritano di essere approfondite separatamente.

In primo luogo, e ciò rappresenta secondo l'opinione del Gruppo di Lavoro *ex* art. 29 un "prerequisito" per ogni trattamento di dati personali[242], il titolare del trattamento è tenuto a "raccogliere" i dati per finalità che siano:

- "*determinate*", ossia specificate sulla base di un'attenta analisi degli scopi per i quali i dati personali verranno utilizzati, escludendo la raccolta di dati non necessari o superflui, in applicazione del principio di necessità ed adeguatezza. Le finalità devono essere chiaramente identificate, ossia deve essere sufficientemente dettagliato quale tipo di trattamento sia incluso ed entro quali limiti le attività del titolare possano ritenersi ricomprese negli scopi indicati. Il momento utile per determinare le finalità è strettamente connesso a quello della raccolta e pertanto deve avvenire prima o contestualmente ad essa. Nulla vieta al titolare di raccogliere i dati personali per più finalità, ferma restando l'applicazione del principio in parola ed un'adeguata enunciazione e determinazione di ciascuno scopo[243];

[242] Cfr. *Opinion 03/2013 on purpose limitation* adottata dal Gruppo di Lavoro ex art. 29 il 02/04/2013 (569/13/EN - WP 203).

[243] Se un titolare fornisce servizi differenti (ad esempio, *email*, *social*, caricamento foto e video), si dovrebbe evitare un'eccessiva semplificazione: un certo livello di dettaglio sarà necessario per assicurare che tutti gli scopi siano sufficientemente chiari per gli utenti (*Example* 3 - *Annex* 3 - *Opinion 03/2013*, cit.). Una informativa stratificata è spesso un modo efficacie per veicolare concetti chiave agli interessati in modo coinciso e immediatamente percepibile, rimandando a successivi livelli le ulteriori informazioni a beneficio di coloro che richiedono chiarimenti aggiuntivi. E così si può pensare di

- "*esplicite*", ossia chiaramente indicate, spiegate o espresse in una forma intelligibile, senza il ricorso a espressioni vaghe, dal carattere ambiguo o altrimenti indecifrabile[244], avuto riguardo ai possibili destinatari
- e alle loro peculiarità sul piano culturale o linguistico[245]. Ciò vale ad escludere anche la legittimità di espressioni troppo tecniche e "legalesi" che non contribuiscono in ultima analisi a chiarire ai destinatari quali siano gli scopi perseguiti dal titolare del trattamento. Il requisito in parola rileva non solo nei confronti dei possibili interessati, ma anche delle autorità di controllo che devono poter essere messe nelle condizioni di valutare la rispondenza di un trattamento agli scopi per i quali i dati personali sono stati raccolti. È importante sottolineare che una mancata o imprecisa indicazione dello scopo o

scomporre una finalità in un determinato numero di sub-finalità, con ciò consentendo anche al titolare di verificare l'operato di eventuali soggetti a cui sono delegate certe tipologie di trattamenti e applicare le necessarie misure di protezione dei dati (*Example 9-10-11 - Annex 3 - Opinion 03/2013*, cit.).

[244] Finalità vaghe o generiche come "*migliorare l'esperienza degli utenti*", "*marketing*", "*IT-security*" o "*ricerche future*", in difetto di ulteriori dettagli, non sodisfano il requisito in parola.
In situazioni dove le finalità possono essere chiaramente ricavate dal contesto è solitamente richiesto un minore grado di dettaglio, salvi i casi in cui possono sorgere ambiguità. Ad esempio: un piccolo commerciante locale viene contattato per consegnare e installare un sistema di riscaldamento presso l'acquirente e per fornire una manutenzione annuale. A tal fine raccoglie informazioni come il nome, l'indirizzo e il numero di telefono del cliente in modo da consegnare e installare il sistema e programmare la revisione annuale. In tale caso gli scopi del trattamento possono essere ricavati implicitamente dal contesto, dalla consuetudine e dalla natura dell'operazione economica sottostante. Tuttavia, se sorgono ambiguità, per esempio, se l'azienda intende inviare anche pubblicità al cliente in merito ad altri suoi servizi (o a servizi di altre aziende), questo dev'essere specificatamente comunicato all'interessato (*Example 5 - Annex 3 - Opinion 03/2013*, cit.).
Ancora. Una piccola ma esclusiva *boutique* specializzata in "vestiti su misura e accessori unici" è solita basare la propria strategia pubblicitaria sul passaparola. L'unico strumento di *marketing* diretto che utilizza è un catalogo annuale che arriva alle case dei 200 clienti in forma cartacea. Una volta iscritti alla lista dei destinatari del catalogo (e come chiaramente indicato nel catalogo stesso), i clienti sono informati della possibilità di annullare l'iscrizione alla *mailing list* in qualunque momento (di persona, per posta, via *email*, o chiamando il negozio). Sono inoltre avvisati che i loro dati non saranno condivisi con altri e saranno utilizzati unicamente per spedire il catalogo. Gli scopi sono sufficientemente specificati in questo semplice contesto. Di contro per una grande azienda di vendita al dettaglio che vende beni tramite il sito Internet in tutta l'Europa e usa *analitycs* complessi per personalizzare le offerte e fare pubblicità mirata, le finalità dovranno essere specificate con molti più dettagli e in modo completo, incluso, fra le altre cose, il modo in cui i dati personali sono trattati. I criteri decisionali utilizzati per il profilare il consumatore devono inoltre essere comunicati (*Example 7-8 - Annex 3 - Opinion 03/2013*, cit.).

[245] Il livello di dettaglio di un'informativa può variare e dev'essere adattato alla fattispecie concreta: un social *network* operante in tutta Europa dovrà prestare particolare attenzione al modo in cui specifica i propri scopi e alla chiarezza delle informazioni fornite, in quanto si rivolge a un ampio gruppo di utenti di diverse culture (*Example 2 - Annex 3 - Opinion 03/2013*, cit.); un sito *web* istituzionale che fornisce consulenze agli anziani o ai malati di mente, un sito di *gaming* rivolto a *teenager*, un'agenzia governativa che elabora tutti i dati dei richiedenti asilo, devono prendere in considerazione l'età, i bisogni speciali, la nazionalità e la cultura degli individui a cui si rivolgono (*Example 4 - Annex 3 - Opinion 03/2013*, cit.).

degli scopi per i quali i dati vengono trattati non autorizza il titolare ad elaborare gli stessi per qualsiasi finalità a propria discrezione o determinarle liberamente sulla base di una interpretazione unilaterale o soggettiva;

- "legittime", ossia devono rispettare tutte le disposizioni in materia di protezione di dati personali ed ogni altra norma applicabile al caso concreto (ad es. diritto del lavoro, disciplina a tutela del consumatore …).

Valutate dunque le caratteristiche che devono possedere le finalità per le quali i dati vengono raccolti, il principio di limitazione dagli scopi prevede che detti dati siano "successivamente trattati in modo che non sia incompatibile con tali finalità". In altre parole, a seguito dell'ottenimento e/o della comunicazione dei dati al titolare del trattamento, occorre che ogni successiva operazione rientri in quell'alveo di finalità che *a priori* è stato individuato ed esplicitato dal titolare stesso. La regola sopra enunciata introduce a tale riguardo un requisito specifico, parlando di compatibilità (o meglio, per usare la doppia negazione dell'art. 5 del GDPR, "non incompatibilità"), che ora è bene delimitare alla luce delle indicazioni dapprima formulate dal Gruppo di Lavoro ex art. 29 ed ora codificate nel nuovo Regolamento.

Innanzitutto, per valutare la compatibilità (o non incompatibilità) con le finalità determinate dal titolare, gli approcci indicati dal Gruppo di Lavoro *ex* art. 29 possono essere di due tipi: formale (ossia mediante confronto tra le finalità inizialmente indicate dal titolare, di solito per iscritto, e qualsiasi ulteriore operazione di trattamento per valutare se sia o meno ricompresa esplicitamente o implicitamente in quanto dichiarato) o sostanziale (ossia mediante una valutazione che prescinda dal dato letterale e prenda in considerazione anche il contesto in cui una certa informativa sia stata resa ed ogni fattore in qualsiasi modo rilevante per comprendere le reali intenzioni del titolare con riguardo al trattamento). Mentre il primo metodo a prima vista può sembrare più obiettivo e neutrale, il rischio di una sua rigida applicazione potrebbe portare a formulazioni di informative molto complesse e per certi versi troppo "aperte", che garantiscano margini ampi per ulteriori operazioni di trattamento che diversamente non troverebbero una giustificazione. Il secondo metodo è più flessibile e pragmatico, e sotto determinati aspetti più efficace e preferibile, consentendo un'interpretazione dinamica adattabile al caso concreto e alle sue evoluzioni, pur senza limitare il diritto ad una protezione efficace dei dati personali.

In taluni casi la compatibilità sarà *prima facie* ovvia e implicita (ad es. a seguito della raccolta iniziale dei dati di un cliente, quali l'indirizzo e le coordinate bancarie, per la consegna periodica di prodotti alimentari a domicilio, è sottinteso che il titolare utilizzerà queste informazioni per organizzare le varie consegne e gestire i singoli pagamenti di volta in volta). In altri casi la stessa compatibilità andrà palesemente esclusa (nello stesso esempio, l'utilizzo di informazioni relative al terminale impiegato dall'utente per inserire gli ordini al fine di proporre sconti personalizzati sulla base del sistema operativo impiegato per gli acquisti, ove non adeguatamente dichiarato nell'informativa, potrà costituire un'operazione di trattamento non compatibile con le finalità per cui si sono raccolti i dati del cliente). Vi è tuttavia una ampia casistica in cui il giudizio di compatibilità si impone come necessario ed è pertanto dirimente individuare quali siano i criteri che consentano di valutare se e in che limiti una finalità possa risultare "non incompatibile".

La norma di cui all'art. 6(4) del GDPR affronta ora in maniera puntuale il problema[246] e identifica quali siano, a titolo esemplificativo ("… tra l'altro …"), gli elementi da

[246] Sul punto si veda anche il Considerando 50 in base al quale "Il trattamento dei dati personali per finalità diverse da quelle per le quali i dati personali sono stati inizialmente raccolti dovrebbe essere consentito solo se compatibile con le finalità per le quali i dati personali sono stati inizialmente raccolti. In tal caso non è richiesta alcuna base giuridica separata oltre a quella che ha consentito la raccolta dei dati personali. Se il trattamento è necessario per l'esecuzione di un compito di interesse pubblico o per l'esercizio di pubblici poteri di cui è investito il titolare del trattamento, il diritto dell'Unione o degli Stati membri può stabilire e precisare le finalità e i compiti per i quali l'ulteriore trattamento è considerato lecito e compatibile. L'ulteriore trattamento a fini di archiviazione nel pubblico interesse, o di ricerca scientifica o storica o a fini statistici dovrebbe essere considerato un trattamento lecito e compatibile. La base giuridica fornita dal diritto dell'Unione o degli Stati membri per il trattamento dei dati personali può anche costituire una base giuridica per l'ulteriore trattamento. Per accertare se la finalità di un ulteriore trattamento sia compatibile con la finalità per la quale i dati personali sono stati inizialmente raccolti, il titolare del trattamento dovrebbe, dopo aver soddisfatto tutti i requisiti per la liceità del trattamento originario, tener conto tra l'altro di ogni nesso tra tali finalità e le finalità dell'ulteriore trattamento previsto, del contesto in cui i dati personali sono stati raccolti, in particolare le ragionevoli aspettative dell'interessato in base alla sua relazione con il titolare del trattamento con riguardo al loro ulteriore utilizzo; della natura dei dati personali; delle conseguenze dell'ulteriore trattamento previsto per gli interessati; e dell'esistenza di garanzie adeguate sia nel trattamento originario sia nell'ulteriore trattamento previsto.
Ove l'interessato abbia prestato il suo consenso o il trattamento si basi sul diritto dell'Unione o degli Stati membri che costituisce una misura necessaria e proporzionata in una società democratica per salvaguardare, in particolare, importanti obiettivi di interesse pubblico generale, il titolare del trattamento dovrebbe poter sottoporre i dati personali a ulteriore trattamento a prescindere dalla compatibilità delle finalità. In ogni caso, dovrebbe essere garantita l'applicazione dei principi stabiliti dal presente regolamento, in particolare l'obbligo di informare l'interessato di tali altre finalità e dei suoi diritti, compreso il diritto di opporsi. L'indicazione da parte del titolare del trattamento di possibili reati o minacce alla sicurezza pubblica e la trasmissione dei dati personali pertinenti a un'autorità competente in singoli casi o in più casi riguardanti lo stesso reato o la stessa minaccia alla sicurezza pubblica dovrebbero essere considerate nell'interesse legittimo perseguito dal titolare del trattamento. Tuttavia, tale trasmissione nell'interesse legittimo del titolare del trattamento o l'ulteriore trattamento

tenere in considerazione per verificare se il trattamento per un'altra finalità sia compatibile con gli scopi per i quali i dati personali sono stati inizialmente raccolti, indicando i seguenti:

- ogni nesso tra le finalità per cui i dati personali sono stati raccolti e le finalità dell'ulteriore trattamento previsto;
- il contesto in cui i dati personali sono stati raccolti, e in particolare le ragionevoli aspettative dell'interessato in base alla sua relazione con il titolare del trattamento con riguardo al loro ulteriore utilizzo;
- la natura dei dati personali (specialmente per particolari categorie di dati personali, oppure se siano trattati dati relativi a condanne penali e a reati);
- le possibili conseguenze dell'ulteriore trattamento previsto per gli interessati;
- l'esistenza di garanzie adeguate sia nel trattamento originario sia nell'ulteriore trattamento previsto, che possono comprendere la cifratura o la pseudonimizzazione, in ogni caso volte ad impedire qualsiasi indebito pregiudizio per l'interessato.

La valutazione di cui si sono appena descritti gli elementi essenziali trova terreno fertile laddove vengano in gioco *big personal data*, rispetto ai quali il Gruppo di Lavoro *ex* art. 29, già nel 2013, segnalava un'importante distinzione nell'approccio al problema con riferimento alle tutele per gli interessati, distinguendo tra uno scenario in cui i titolari elaborano i dati per rilevare generalmente le tendenze e le correlazioni tra le informazioni e uno in cui l'interesse è focalizzato sull'individuo[247].

E questa distinzione e differente impostazione trova peraltro oggi il suo connotato normativo nell'art. 5(1)(b) del GDPR sopra richiamato che, a prescindere da un giudizio di compatibilità, considera "non incompatibile" con le finalità iniziali "un ulteriore trattamento dei dati personali a fini di archiviazione nel pubblico interesse, di ricerca scientifica o storica o a fini statistici".

In conclusione, il trattamento per scopi diversi da quelli per i quali i dati sono stati raccolti non è di per sé illegittimo, ma, laddove non avvenga per fini di archiviazione nel pubblico interesse, di ricerca scientifica o storica o a fini statistici, o la sua

dei dati personali dovrebbero essere vietati se il trattamento non è compatibile con un obbligo vincolante di segretezza, di natura giuridica, professionale o di altro genere".

[247] Cfr. *Opinion 03/2013*, cit., *Annex 2*, laddove viene avanzata la domanda "*what safeguards would make the further use of personal data for analytics compatible?*".

giustificazione non risieda nel consenso dell'interessato ovvero su un atto legislativo dell'Unione o degli Stati membri che costituisca una misura necessaria e proporzionata in una società democratica per la salvaguardia degli obiettivi di salvaguardia della sicurezza pubblica[248], esso richiede una attenta ed oculata valutazione da svolgersi caso per caso secondo le peculiarità della fattispecie in esame sulla base del principio di limitazione delle finalità qui analizzato.

8. Le nuove sfide poste dai Big Data

8.1 I rischi dal punto di vista *privacy* e il nuovo approccio

È opinione condivisa che un responsabile approccio ai Big Data (ed in particolare ai *big personal data*) possa riservare notevoli benefici e maggiore efficienza alla società e alle persone nei più disparati campi, dalla ricerca all'ambiente, dall'energia alla salute.

Non mancano tuttavia preoccupazioni sull'impatto che un utilizzo massivo di informazioni, anche avente carattere personale, ha (o può avere) sui diritti e le libertà delle persone, compreso il loro diritto alla vita privata. Si pensi ai modelli commerciali che profilano i soggetti in base alle loro propensioni di acquisto e di consumo, alla circolazione in tempo reale dei dati, alla moltiplicazione delle possibili finalità a cui sono volti i trattamenti. Tutte queste nuove pressioni, laddove coinvolgano direttamente le persone e la loro sfera di libertà, meritano oggi di essere inquadrate e risolte nell'ambito della normativa *privacy* secondo i principi da sempre noti in materia, da un lato, e, per altro verso, alla luce dei dettami del GDPR che, come si metterà brevemente in luce, offre una serie di strumenti per affrontare le nuove sfide poste dai Big Data nella prospettiva di una costante tutela dei soggetti coinvolti.

Quanto verrà esposto prende le mosse in particolare dal parere *Meeting the Challenges of Big Data*[249] con cui il Garante Europeo della Protezione dei Dati ha dichiarato di essere intenzionato ad avviare un nuovo dibattito aperto con i legislatori, le autorità di controllo, i rappresentanti del settore, gli esperti del settore informatico, il mondo scientifico ed accademico e la società civile per capire come sfruttare i benefici sociali

[248] Cfr. art. 23 del GDPR.
[249] Opinion 7/2015, cit.

offerti dai Big Data tutelando al tempo stesso la dignità e i diritti e le libertà fondamentali degli individui in modo più efficace e innovativo.

"L'uso dei Big Data richiede una maggiore protezione dei dati; sarà quindi indispensabile un maggiore controllo sugli utilizzatori di questi dati per garantirne un impiego responsabile in futuro. Norme sulla *privacy* sono state definite proprio a tutela dei nostri diritti e valori fondamentali. La domanda che l'industria e gli enti pubblici devono porsi non è se applicare o meno tali norme al trattamento dei Big Data bensì come applicarle più efficacemente. Siamo intenzionati a collaborare con tutti gli interlocutori chiave, al di fuori e all'interno dell'UE, al fine di individuare soluzioni creative e orientate al futuro per preservare al meglio i nostri valori, senza con ciò rinunciare ai benefici sociali nell'interesse pubblico"[250].

Questi gli elementi fondamentali sui quali dovrebbe basarsi uno sviluppo "responsabile e sostenibile" dei Big Data nell'ottica del Garante Europeo della Protezione dei Dati:

- assicurare maggiore trasparenza riguardo al modo in cui si trattano i dati personali (*trasparency*);
- garantire agli utilizzatori un maggiore controllo sul modo in cui i loro dati vengono utilizzati (*user control*);
- integrare nei prodotti e servizi una protezione dei dati facilmente comprensibile fin dalla fase di progettazione (*privacy by design*); e
- responsabilizzarsi ulteriormente rispetto a quello che fanno (*accountability*).

8.2 Trasparenza

In tema di trasparenza ciò che viene ampiamente sottolineato dal Garante Europeo della Protezione dei Dati concerne la necessità che gli interessati ricevano informazioni chiare su quali dati a loro riferiti vengono trattati, per quali finalità e con quali modalità, ivi incluse le logiche utilizzate negli algoritmi per determinare presunzioni o ipotesi che riguardano gli interessati stessi.

"*Transparency of automated decisions is taking an increasingly important role with the advent of Big Data analytics. Disclosing the logic of decision-making can help individuals better to verify whether the conclusions drawn by the organisations*

[250] Comunicato Stampa EDPS/2015/11 Bruxelles, 19/11/2015.

processing the data and impacting the individuals are accurate and fair. They can better understand, and perhaps rectify, the criteria underpinning, and the factors influencing the decision"[251].

La questione in oggetto ha trovato diverse declinazioni nel GDPR: la parola "trasparenza" non solo si impone come principio cardine che governa la materia (l'art. 5(1)(a) prescrive che i dati personali siano trattati "in modo lecito, corretto e trasparente nei confronti dell'interessato"), ma si ritrova in numerosi passaggi della menzionata normativa[252] ed in particolare alla Sezione I del Capo III ("Trasparenza e modalità"), laddove viene imposto al titolare di adottare misure appropriate per fornire all'interessato tutte le informazioni relative al trattamento "in forma concisa, trasparente, intelligibile e facilmente accessibile" (art. 12).

---

[251] *Opinion 7/2015*, cit.

[252] Si vedano il Considerando 39 ("[…] Dovrebbero essere trasparenti per le persone fisiche le modalità con cui sono raccolti, utilizzati, consultati o altrimenti trattati dati personali che li riguardano nonché la misura in cui i dati personali sono o saranno trattati. Il principio della trasparenza impone che le informazioni e le comunicazioni relative al trattamento di tali dati personali siano facilmente accessibili e comprensibili e che sia utilizzato un linguaggio semplice e chiaro. Tale principio riguarda, in particolare, l'informazione degli interessati sull'identità del titolare del trattamento e sulle finalità del trattamento e ulteriori informazioni per assicurare un trattamento corretto e trasparente con riguardo alle persone fisiche interessate e ai loro diritti di ottenere conferma e comunicazione di un trattamento di dati personali che li riguardano. È opportuno che le persone fisiche siano sensibilizzate ai rischi, alle norme, alle garanzie e ai diritti relativi al trattamento dei dati personali, nonché alle modalità di esercizio dei loro diritti relativi a tale trattamento. In particolare, le finalità specifiche del trattamento dei dati personali dovrebbero essere esplicite e legittime e precisate al momento della raccolta di detti dati personali […]"), il Considerando 58 ("Il principio della trasparenza impone che le informazioni destinate al pubblico o all'interessato siano concise, facilmente accessibili e di facile comprensione e che sia usato un linguaggio semplice e chiaro, oltre che, se del caso, una visualizzazione. Tali informazioni potrebbero essere fornite in formato elettronico, ad esempio, se destinate al pubblico, attraverso un sito web […]"), il Considerando 60 ("I principi di trattamento corretto e trasparente implicano che l'interessato sia informato dell'esistenza del trattamento e delle sue finalità. Il titolare del trattamento dovrebbe fornire all'interessato eventuali ulteriori informazioni necessarie ad assicurare un trattamento corretto e trasparente, prendendo in considerazione le circostanze e del contesto specifici in cui i dati personali sono trattati"), il Considerando 71, specificamente rivolto all'attività di profilazione ("[…] Al fine di garantire un trattamento corretto e trasparente nel rispetto dell'interessato, tenendo in considerazione le circostanze e il contesto specifici in cui i dati personali sono trattati, è opportuno che il titolare del trattamento utilizzi procedure matematiche o statistiche appropriate per la profilazione, metta in atto misure tecniche e organizzative adeguate al fine di garantire, in particolare, che siano rettificati i fattori che comportano inesattezze dei dati e sia minimizzato il rischio di errori e al fine di garantire la sicurezza dei dati personali secondo una modalità che tenga conto dei potenziali rischi esistenti per gli interessi e i diritti dell'interessato e che impedisca tra l'altro effetti discriminatori nei confronti di persone fisiche sulla base della razza o dell'origine etnica, delle opinioni politiche, della religione o delle convinzioni personali, dell'appartenenza sindacale, dello status genetico, dello stato di salute o dell'orientamento sessuale, ovvero che comportano misure aventi tali effetti […]"), il Considerando 100 ("Al fine di migliorare la trasparenza e il rispetto del presente regolamento dovrebbe essere incoraggiata l'istituzione di meccanismi di certificazione e sigilli nonché marchi di protezione dei dati che consentano agli interessati di valutare rapidamente il livello di protezione dei dati dei relativi prodotti e servizi").

Senza entrare nel dettaglio, si rimanda agli artt. 13-14 per quanto riguarda il contenuto dell'informativa, richiamando tuttavia l'attenzione su uno degli elementi che, nell'analizzare le problematiche connesse ai possibili impieghi di *big personal data*, si impone maggiormente per la sua possibile portata innovativa rispetto alla legislazione previgente: si tratta della necessaria informazione resa all'interessato circa la "esistenza di un processo decisionale automatizzato, compresa la profilazione […], e, almeno in tali casi, informazioni significative sulla logica utilizzata, nonché l'importanza e le conseguenze previste di tale trattamento per l'interessato"[253].

8.3 *User control* e portabilità

Strettamente connessa al tema della trasparenza, nell'ottica del legislatore comunitario vi è la questione del controllo da parte dell'interessato sui trattamenti di dati che lo riguardano: solo attraverso un effettivo potere degli interessati è possibile garantire l'attuazione e il rispetto di pratiche leali e corrette e un efficace contrasto agli errori. Ciò, in primo luogo, come visto, sui possibili usi secondari dei dati, resi ancor più facili dai Big Data e dalle tecnologie che su essi si basano. Ma anche con riguardo a tutte le operazioni in qualsiasi modo connesse con la privacy degli individui che con il proliferare del numero di dati trattati accrescono le preoccupazioni e, conseguentemente, le cautele da osservare.

La tematica presenta due distinte sfaccettature che di seguito si riassumono. In prima analisi il controllo dell'interessato si presenta come una sua legittima ingerenza nel trattamento dei dati a lui riferiti. Emblematica sotto questo aspetto è la formulazione dell'art. 13(2) del GDPR (ma specularmente anche dell'art. 14(2) con riguardo alle informazioni da fornire qualora i dati personali non siano stati ottenuti presso l'interessato[254]), secondo cui il titolare del trattamento ha l'obbligo di informare

[253] Art. 13(2)(f) e, del tutto specularmente, per quanto riguarda le informazioni da fornire qualora i dati non siano stati ottenuti presso l'interessato, l'art. 14(2)(g).

[254] Sul punto si vedano anche il Considerando 59 ("È opportuno prevedere modalità volte ad agevolare l'esercizio, da parte dell'interessato, dei diritti di cui al presente regolamento, compresi i meccanismi per richiedere e, se del caso, ottenere gratuitamente, in particolare l'accesso ai dati, la loro rettifica e cancellazione e per esercitare il diritto di opposizione. Il titolare del trattamento dovrebbe predisporre anche i mezzi per inoltrare le richieste per via elettronica, in particolare qualora i dati personali siano trattati con mezzi elettronici […]"), il Considerando 63 ("Un interessato dovrebbe avere il diritto di accedere ai dati personali raccolti che la riguardano e di esercitare tale diritto facilmente e a intervalli ragionevoli, per essere consapevole del trattamento e verificarne la liceità […]. Ogni interessato dovrebbe pertanto avere il diritto di conoscere e ottenere comunicazioni in particolare in relazione alla finalità per cui i dati personali sono trattati, ove possibile al periodo in cui i dati personali sono trattati,

l'interessato dei seguenti diritti (poi meglio sviluppati nel seguito della normativa), aventi ad oggetto:

- l'accesso ai dati personali (art. 15): su questo punto si concentrano le maggiori attenzioni da parte del Garante Europeo che, analizzando le sfide poste dai Big Data, rileva che "*the right to access and correct one's personal data is one of the fundamental principles of European data protection law and is becoming increasingly more important with advances of Big Data analytics. Individuals must be empowered to better detect unfair biases and challenge mistakes arising from the logic used in algorithms to determine assumptions and predictions and a strong right of access and correction is a precondition to this*"[255]. Il tema introduce e sollecita anche ulteriori riflessioni riguardanti la conservazione e l'archiviazione dei dati e la possibilità di utilizzare per tali finalità piattaforme c.d. *cloud*;
- la rettifica, senza ingiustificato ritardo, dei dati personali inesatti (art. 16);
- la cancellazione dei dati personali (art. 17)[256];

---

ai destinatari dei dati personali, alla logica cui risponde qualsiasi trattamento automatizzato dei dati e, almeno quando è basato sulla profilazione, alle possibili conseguenze di tale trattamento. Ove possibile, il titolare del trattamento dovrebbe poter fornire l'accesso remoto a un sistema sicuro che consenta all'interessato di consultare direttamente i propri dati personali"), il Considerando 70 in tema di marketing ("Qualora i dati personali siano trattati per finalità di marketing diretto, l'interessato dovrebbe avere il diritto, in qualsiasi momento e gratuitamente, di opporsi a tale trattamento, sia con riguardo a quello iniziale o ulteriore, compresa la profilazione nella misura in cui sia connessa a tale marketing diretto. Tale diritto dovrebbe essere esplicitamente portato all'attenzione dell'interessato e presentato chiaramente e separatamente da qualsiasi altra informazione") e il Considerando 71 in tema di profilazione ("L'interessato dovrebbe avere il diritto di non essere sottoposto a una decisione, che possa includere una misura, che valuti aspetti personali che lo riguardano, che sia basata unicamente su un trattamento automatizzato e che produca effetti giuridici che lo riguardano o incida in modo analogo significativamente sulla sua persona, quali il rifiuto automatico di una domanda di credito online o pratiche di assunzione elettronica senza interventi umani […]").

[255] *Opinion 7/2015*, cit.

[256] Una delle novità più rilevanti del GDPR, codificata all'art. 17, riguarda il c.d. diritto all'oblio, in forza del quale l'interessato ha il diritto di ottenere dal titolare del trattamento (che conseguentemente ha l'obbligo di adempiere "senza giustificato ritardo" e salvi i casi previsti di trattamento necessario per taluni scopi) la cancellazione dei dati personali che lo riguardano al ricorrere di uno dei seguenti motivi:

a) i dati personali non sono più necessari per gli scopi per i quali sono stati raccolti;
b) l'interessato revoca il consenso su cui si basa il trattamento (salvo altro fondamento giuridico);
c) l'interessato si oppone legittimamente al trattamento nelle forme previste;
d) i dati personali sono stati trattati illecitamente;
e) i dati personali devono essere cancellati per adempiere un obbligo legale;
f) i dati personali relativi a minori sono stati raccolti nell'ambito dell'offerta di servizi della società dell'informazione.

Per rafforzare il diritto all'oblio nell'ambiente online, il titolare ha l'obbligo di comunicare la richiesta di cancellazione a chiunque li stia trattando - nei limiti di quanto tecnicamente possibile - chiedendo di "cancellare qualsiasi link verso tali dati personali o copia o riproduzione di detti dati personali" (cfr. sul punto anche Considerando 66).

- la limitazione[257] del trattamento (art. 18);
- l'opposizione per motivi connessi alla situazione particolare dell'interessato (art. 21), ovvero a fronte di decisioni basate unicamente sul trattamento automatizzato, compresa la profilazione, che producano effetti giuridici che riguardino l'interessato o che incidano in modo analogo e significativamente sulla sua persona (art. 22);
- la revoca del consenso in qualsiasi momento, senza pregiudicare la liceità del trattamento basata sul consenso precedentemente prestato prima della revoca;
- la possibilità di proporre reclamo a un'autorità di controllo.

Tuttavia, il concetto di "controllo" non si manifesta soltanto attraverso strumenti di tutela (siano essi di accesso ai dati o di reazione a uno o più determinati trattamenti), ma viene ulteriormente rafforzato mettendo al servizio dell'interessato alcuni dei benefici e delle potenzialità delle nuove tecnologie che si basano sui Big Data, contribuendo così ad un approccio ancor più efficiente e trasparente al tema della *privacy*. A questo proposito il riferimento va ai concetti di portabilità e interoperabilità, in base ai quali la normativa comunitaria (art. 20 del GDPR[258]) assicura oggi a ciascun interessato il diritto di:

- ricevere un sottoinsieme di dati personali che lo riguardano;
- trasmettere "senza impedimenti" dati personali a un altro titolare del trattamento. Tale trasmissione, "se tecnicamente fattibile", può essere richiesta in via "diretta" da un titolare all'altro (appartenente allo stesso o a un diverso settore di attività). In questo senso, mentre il titolare trasmittente non ha alcun

[257] Limitazione che l'interessato ha il diritto di ottenere al ricorrere di una delle seguenti ipotesi: a) qualora l'interessato contesti l'esattezza dei dati personali, per il periodo necessario al titolare del trattamento per verificare l'esattezza di tali dati personali; b) qualora il trattamento sia illecito e l'interessato si opponga alla cancellazione dei dati personali e chieda invece che ne sia limitato l'utilizzo; c) benché il titolare del trattamento non ne abbia più bisogno ai fini del trattamento, qualora i dati personali siano necessari all'interessato per l'accertamento, l'esercizio o la difesa di un diritto in sede giudiziaria; d) qualora l'interessato si sia opposto al trattamento, in attesa della verifica in merito all'eventuale prevalenza dei motivi legittimi del titolare del trattamento rispetto a quelli dell'interessato.

[258] Sul punto si veda anche il Considerando 68 in forza del quale "Per rafforzare ulteriormente il controllo sui propri dati è opportuno anche che l'interessato abbia il diritto, qualora i dati personali siano trattati con mezzi automatizzati, di ricevere in un formato strutturato, di uso comune, leggibile da dispositivo automatico e interoperabile i dati personali che lo riguardano che abbia fornito a un titolare del trattamento e di trasmetterli a un altro titolare del trattamento. È opportuno incoraggiare i titolari del trattamento a sviluppare formati interoperabili che consentano la portabilità dei dati […]". Sempre in tema cfr. Linee guida sul diritto alla "portabilità dei dati", adottate dal Gruppo di Lavoro ex art. 29 il 13/12/2016 - versione emendata e adottata il 05/04/2017.

obbligo specifico di verificare la qualità dei dati prima di trasmetterli, il titolare ricevente è tenuto a garantire che i dati ricevuti siano pertinenti e non eccedenti rispetto al nuovo trattamento svolto[259].

Il diritto alla portabilità, oltre a garantire la libera circolazione dei dati personali nell'Unione Europea e favorire la concorrenza fra i titolari, è strutturato in modo da non pregiudicare nessuno degli altri diritti previsti in capo all'interessato e trova applicazione al simultaneo ricorrere dei seguenti due requisiti:

a) il trattamento si basi sul consenso o su un contratto;
b) il trattamento sia effettuato con mezzi automatizzati.

L'art. 12 del GDPR dispone che la richiesta dell'interessato venga gestita "senza ingiustificato ritardo" e comunque "entro un mese" (par. 3) e vieta al titolare di addebitare oneri all'interessato per la fornitura dei dati personali, salvo dimostrare il carattere manifestamente infondato o eccessivo delle richieste "in particolare per il loro carattere ripetitivo" (par. 5).

Quanto infine al formato, la menzionata normativa *privacy* pone in capo ai titolari di trattamento l'obbligo di fornire i dati personali richiesti dall'interessato in modo da consentirne il riutilizzo (testualmente all'art. 20 viene detto "in un formato strutturato, di uso comune e leggibile da dispositivo automatico"). Nel Considerando 68 del Regolamento in commento (già citato in nota) si chiarisce ulteriormente che il formato in questione dovrebbe essere "interoperabile". Per una definizione normativa a livello europeo di interoperabilità si richiama l'art. 2 della Decisione n. 922/2009/CE del Parlamento europeo e del Consiglio, del 16/09/2009, relativa a soluzioni interoperabili per le amministrazioni pubbliche europee (ISA) che utilizza la seguente espressione: "capacità di organizzazioni diverse e disparate di interagire in vista di obiettivi comuni concordati e reciprocamente vantaggiosi, ricorrendo alla condivisione di conoscenze e informazioni tra le organizzazioni, per mezzo dei processi aziendali che su di esse si basano, tramite lo scambio di dati fra i rispettivi sistemi TIC".

[259] A titolo esemplificativo si consideri che in caso di una richiesta di portabilità rivolta a un servizio di posta elettronica via *web*, se la richiesta serve all'interessato per recuperare i messaggi di posta elettronica inviandoli a una piattaforma di archiviazione, il nuovo titolare non ha necessità di trattare le informazioni di contatto dei soggetti con cui l'interessato ha scambiato messaggi. Se le informazioni non sono pertinenti rispetto alle finalità del nuovo trattamento, allora non devono essere conservate o trattate (Linee guida sul diritto alla "portabilità dei dati", adottate dal Gruppo di Lavoro *ex* art. 29 il 13/12/2016 - versione emendata e adottata il 05/04/2017).

Ciò che appariva (e può apparire) come un problema o un ostacolo all'effettivo rispetto della sfera personale e della dignità dell'individuo, può dunque tradursi in una potenzialità, non solo in termini competitivi per le imprese, ma anche rispetto ai soggetti a cui sono riferiti i dati oggetto di trattamento, i quali sono messi nelle condizioni di beneficiare delle possibilità offerte dalla *data science*, come ben messo in luce dal Garante Europeo della Protezione dei Dati: "*Allowing data portability could enable businesses and individuals to minimize the benefits of Big Data in a more balanced and transparent way and may help redress the economic imbalance between controllers on one hand and individuals on the other. It could also let individuals benefit from the value created by the use of their personal data: it could allow them to use the data for their own purposes, or to license the data for further use to third parties, in exchange of additional services, or for cash value. Further, it could also help to minimize unfair or discriminatory practices and reduce the risks of using inaccurate data for decision-making purposes*"[260].

8.4 *Privacy by default* e *privacy by design*

In risposta alle sfide poste dai Big Data, il Garante Europeo della Protezione dei Dati ha osservato che gli attuali capisaldi in materia di *privacy*, quali la necessità, la proporzionalità, la minimizzazione dei dati, la limitazione delle finalità e la trasparenza, devono rimanere principi chiave e fornire una "linea di base necessaria per proteggere i nostri diritti fondamentali in un mondo di megadati"[261].

Tali principi, al contempo e in ragione proprio delle nuove realtà tecnologiche con cui il mondo attuale si confronta, devono essere rafforzati e applicati con maggiore efficacia e "in modo più moderno, flessibile, creativo e innovativo".

Per venire incontro a tali esigenze il GDPR all'art. 25 introduce i concetti di *privacy by default* (ossia protezione dei dati per impostazione predefinita) e *privacy by design* (ossia protezione dei dati fin dalla progettazione)[262].

---

[260] *Opinion 7/2015*, cit.

[261] Garante Europeo della Protezione dei Dati (19/11/2015) - Sintesi esecutiva del parere del Garante europeo della protezione dei dati: «La risposta alle sfide dei megadati: richiesta di trasparenza, controllo da parte degli utilizzatori, protezione dei dati fin dalla progettazione e responsabilità».

[262] Si veda sul punto anche il Considerando 78 in base al quale "La tutela dei diritti e delle libertà delle persone fisiche relativamente al trattamento dei dati personali richiede l'adozione di misure tecniche e organizzative adeguate per garantire il rispetto delle disposizioni del presente regolamento. Al fine di poter dimostrare la conformità con il presente regolamento, il titolare del trattamento dovrebbe adottare

Con il primo concetto viene stabilito che, "per impostazione predefinita", i titolari dovrebbero trattare solo i dati personali nella misura necessaria per le finalità previste e per il periodo strettamente necessario a tali fini, in particolare non rendendo accessibili dati personali a un numero indefinito di persone fisiche.

Con il secondo si attribuisce al titolare un ruolo maggiormente attivo (e proattivo), coinvolgendolo e incoraggiandolo ad affrontare le tematiche connesse alla *privacy* in ogni fase di sviluppo, progettazione, selezione e utilizzo di applicazioni, servizi e prodotti basati sul trattamento di dati personali. Ciò tenuto conto "dello stato dell'arte e dei costi di attuazione, nonché della natura, dell'ambito di applicazione, del contesto e delle finalità del trattamento, come anche dei rischi aventi probabilità e gravità diverse per i diritti e le libertà delle persone fisiche costituiti dal trattamento" (art. 25). Proprio sotto questo aspetto, il Garante Europeo della Protezione dei Dati ritiene fondamentale per uno sviluppo responsabile e sostenibile dei Big Data, integrare nei prodotti e servizi una protezione dei dati "facilmente comprensibile fin dalla fase di progettazione", sfruttando le potenzialità delle nuove tecnologie, a partire da idonee ed efficaci interfacce grafiche con cui permettere un dialogo diretto ed immediato tra titolare del trattamento e interessato.

"*Individuals need to be offered new, innovative ways to be informed about what happens to their data, and to exercise control over their data. This requires innovative and privacy-friendly engineering as well as privacy-friendly organisational arrangements and business practices. Innovative and responsible engineering can facilitate, among others, the exercise of individuals' rights of access, objection, opt-out, correction, as well as data portability. Privacy-friendly engineering can also be invaluable in helping develop new business models for generating value from for example, data stores*"[263].

8.5 *Accountability*

Un corretto approccio al tema dei Big Data (anzi, dei *big personal data*) non può in ultima analisi che passare anche da una consapevole responsabilizzazione (in inglese *accountability*) dei titolari del trattamento.

---

politiche interne e attuare misure che soddisfino in particolare i principi della protezione dei dati fin dalla progettazione e della protezione dei dati di *default*".

[263] *Opinion 7/2015*, cit.

Vari sono gli istituti che nel GDPR, realizzano questo obiettivo (a partire dall'ultimo paragrafo dell'art. 5 in tema di principi applicabili al trattamento di personali), ponendo al centro la valutazione dei rischi connessi al trattamento e sollecitando i soggetti coinvolti a orientare la propria struttura e organizzazione (cfr. *supra*, *privacy by design*) verso una maggiore consapevolezza delle problematiche connesse alla *privacy*.

Oltre agli incentivi verso l'elaborazione di codici di condotta (artt. 40-41) e l'istituzione di meccanismi di certificazione della protezione dei dati anche attraverso sigilli e marchi (artt. 42-43), si riportano di seguito alcune novità recentemente introdotte, evidenziando come la loro applicabilità al singolo caso dipenderà non soltanto dalla dimensione del titolare, ma anche dalle problematiche sottese e dai rischi per i diritti e le libertà degli interessati (rischi che l'uso di nuove tecnologie, la natura, l'oggetto, il contesto e le finalità del trattamento possono accrescere):

- registri delle attività di trattamento, da tenersi in forma scritta, anche in formato elettronico, e mettere a disposizione delle autorità di controllo per tutte quelle realtà con almeno 250 dipendenti ovvero che effettuino trattamenti non occasionali che presentino un rischio per i diritti e le libertà dell'interessato o riguardino categorie particolari di dati (art. 31);
- misure tecniche e organizzative da implementare allo scopo di garantire un livello di sicurezza adeguato al rischio, quali, a titolo esemplificativo la pseudonimizzazione e la cifratura dei dati personali; la capacità di assicurare su base permanente la riservatezza, l'integrità, la disponibilità e la resilienza dei sistemi e dei servizi di trattamento; la capacità di ripristinare tempestivamente la disponibilità e l'accesso dei dati personali in caso di incidente fisico o tecnico; una procedura per testare, verificare e valutare regolarmente l'efficacia delle misure tecniche e organizzative al fine di garantire la sicurezza del trattamento (art. 32);
- notifica all'autorità di controllo e comunicazione all'interessato di una violazione dei dati (c.d. *data breach*) da compiersi senza ingiustificato ritardo, e comunque, per quanto riguarda l'autorità, entro 72 ore dal momento dell'avvenuta conoscenza, descrivendo la natura della violazione, le categorie e il numero approssimativo di interessati, le categorie e il numero

approssimativo di registrazioni dei dati personali, le probabili conseguenze, le misure adottate o di cui si propone l'adozione per porre rimedio e anche, se del caso, per attenuarne i possibili effetti negativi (art. 34);

- valutazione d'impatto sulla protezione dei dati (in inglese *Data Protection Impact Assessment*), da compiersi prima di procedere al trattamento nei casi di una valutazione sistematica e globale di aspetti personali relativi a persone fisiche, basata su un trattamento automatizzato, compresa la profilazione, e sulla quale si fondano decisioni che hanno effetti giuridici o incidono in modo analogo significativamente su dette persone fisiche; trattamento, su larga scala, di categorie particolari di dati personali; sorveglianza sistematica su larga scala di una zona accessibile al pubblico (art. 35);
- designazione del Responsabile della Protezione dei Dati (in inglese *Data Protection Officer* o DPO), ossia di un soggetto con conoscenza specialistica della materia, qualità professionale e capacità di assolvere i propri compiti, quale supporto per il titolare in merito agli obblighi *privacy* (artt. 37-39).

Quanto appena riportato vuole essere una rapida panoramica sulle implicazioni giuridiche in tema di *privacy* che il dibattito sui *big* (*personal*) *data* sollecita, imponendo a ciascun soggetto coinvolto un'effettiva consapevolezza delle problematiche sottese e delle cautele da adottare nell'affrontare l'argomento. Il tutto nella certezza di muoversi in un quadro legislativo recentemente rinnovato, che già in parte ha recepito alcune delle preoccupazioni legate ai possibili pregiudizi per la libertà degli individui e che, in un'ottica di costante aggiornamento, cerca di stimolare la formulazione di norme sempre più rigorose sulla tutela della vita privata nel mondo digitale.

Come affermato dal Garante Europeo della Protezione dei dati "le opportunità che offrono i megadati in termini di incremento della produttività e della connettività dovrebbero essere accompagnate da corrispondenti garanzie di tutela. […] La strategia per il mercato unico digitale è per l'UE l'opportunità giusta per lavorare in modo coerente per il conseguimento di tali obiettivi"[264].

[264] Garante Europeo della Protezione dei Dati (23/09/2016) - Sintesi del parere del Garante europeo della protezione dei dati su un'applicazione efficace delle normative nell'economia della società digitale.

"*European data protection law has been developed to protect our fundamental rights and values, including our right to privacy. The question is not whether to apply data protection law to Big Data, but rather how to apply it innovatively in new environments. Our current data protection principles, including transparency, proportionality and purpose limitation, provide the base line we will need to protect more dynamically our fundamental rights in the world of Big Data. They must, however, be complemented by 'new' principles which have developed over the years such as accountability and privacy by design and by default. The EU data protection reform package is expected to strengthen and modernise the regulatory framework*"[265].

[265] *Opinion 7/2015*, cit.

CONCLUSIONI

Il fenomeno dei *Big Data*, probabilmente una rivoluzione equiparabile a quella del digitale e delle reti, sta cambiando i modi di produrre, commercializzare e consumare: in altre parole, le nostre vite.

E lo sta facendo in modo totalmente trasversale: nell'aprile 2017 due giovani ingegneri indiani hanno messo a punto un algoritmo onde avvertire gli automobilisti del pericolo di collisione con mucche vaganti (a quanto pare una delle cause principali di incidenti in India).

Nel campo delle neuroscienze, i Big Data vengono utilizzati per cercare di mappare le connettività neurali del cervello: grazie a queste tecniche si è riusciti a costruire un modello 3D del cervello di un moscerino da frutta, passando al setaccio, in un periodo di dieci anni, *terabyte* di dati di diversa natura e tipologia per ricostruire 1000 cellule nervose. Negli Stati Uniti, un algoritmo viene invece utilizzato per valutare se concedere o meno la libertà su cauzione, sfruttando una valutazione di rischio con modelli predittivi basati su parametri oggettivi elaborati dal programma.

Ma non mancano esempi che svelano l'altra faccia, quella più ambigua ed allarmante del fenomeno. Si ricorderà la notizia della pubblicazione ad opera di tal *Orangita Books*, sedicente centro culturale, che ha pubblicato il "*Catalogo delle donne single di Lecco*", rassegna di tutti i profili *Facebook* in cui erano presenti tre semplici filtri, "città in cui vivi" (Lecco), "situazione sentimentale" (*single*), "genere" (donna).

Di tutt'altro tenore, ma non per questo più rassicuranti, sono le riflessioni sollevate dalla pratica di alcune *start-up* nell'ambito *fintech* che gestiscono piattaforme di prestito *peer-to-peer* e che utilizzano algoritmi in grado di raccogliere ed abbinare dati da svariate fonti (non sempre note all'utente) per costruire tabelle di *credit scoring* utilizzato poi per determinare il tasso del prestito.

Ancor più recentemente, il caso *Cambridge Analytica*, applicazione che ha raccolto innumerevoli dati da test di personalità ed in generale dal comportamento degli utenti di *Facebook* per poi riutilizzarli nelle campagne politiche più importanti dell'ultimo biennio, ha aperto gli occhi sull'influenza che la gestione dei dati può avere nella formazione del consenso.

Insomma, i *Big Data* sono in grado di essere utilizzati praticamente in tutti i settori della vita sociale e sono sempre maggiori le loro applicazioni; e con tali applicazioni, aumentano anche le opportunità e le ambiguità da considerare. Si tratta di un fenomeno ancora molto fluido, di cui non è facile allo stato definire in modo netto i confini e le implicazioni teoriche e pratiche. Tuttavia è un fenomeno sempre più diffuso e riguardante sempre più settori ed attività.

Per comprenderne in profondità la portata, occorre, però, fare un passo indietro ed analizzare, in termini più sintetici ed essenziali possibili, il contesto entro cui il fenomeno dei *Big Data* ha avuto origine: la digitalizzazione.

In tal senso, è opinione comune e diffusa che l'avvento delle tecnologie informatiche e l'invenzione della "rete" web stiano apportando trasformazioni radicali negli stili di vita di ogni giorno, nelle realtà economiche, nelle modalità di impresa e nello stesso essere e concepirsi ciascuno come parte di un mondo che non ha più solo una faccia analogica.

Da più parti si parla di "*digital disruption*", ovvero la rivoluzione digitale che sta ridisegnando radicalmente mercati e settori industriali e che ridetermina i termini di competizione tra le imprese.

Internet e la digitalizzazione in genere hanno portato ad una straordinaria diffusione e democratizzazione di saperi e conoscenze, riequilibrando le asimmetrie informative tra venditori e compratori e contribuendo ad una maggiore trasparenza di mercato: si pensi ai siti di comparazione (*price comparison website*), divenuti ormai imprescindibili per l'acquisto di un biglietto aereo, di un bene o per la prenotazione di una camera d'albergo.

Sono inoltre diminuiti vertiginosamente i tempi ed i costi di ricerca di ciò che si desidera acquistare o conoscere, permettendo funzionalità integrate che hanno scalzato servizi e prodotti introdotti pochi anni prima: oggi con uno *smartphone* si può fare a meno di orologi, sveglie, navigatori, macchine fotografiche, dispositivi per ascoltare la musica, giornali, riviste, guide turistiche, essendo tutto disponibile in un unico dispositivo tascabile.

Dal punto di vista delle imprese, il disporre di conoscenze altamente qualificate in ambito tecnologico ed informatico permette di abbattere barriere all'entrata in settori

che prima sembravano quasi del tutto impenetrabili (si pensi alla rapida espansione inter-settoriale dell'offerta di società come *Amazon* o *Google*).

Inoltre la digitalizzazione impone una continua pressione verso l'innovazione di processi produttivi (*Foodora* o *Deliveroo* che stanno rivoluzionando l'idea stessa di *supply chain*), metodi di business (*Uber* e *Bla Bla Car* per il mercato dei trasporti privati; *Airbnb* per il settore alberghiero) ed offerte (*Netflix* e *Spotify* rispetto al panorama dello sfruttamento online di opere, rispettivamente, cinematografiche e musicali), pena l'esclusione, anche in tempi rapidi, dal mercato (come *Blockbuster* insegna).

Accanto, però, agli evidenti ed innegabili benefici della digitalizzazione, sono sempre crescenti i timori e le perplessità per le criticità che tale rivoluzione mostra: l'impatto sui posti di lavoro che sembrano destinati inevitabilmente a diminuire, sostituiti sempre più da robot e processi automatizzati; i timori sulla segretezza e riservatezza delle informazioni veicolate in qualche modo in rete e sulla vulnerabilità rispetto ad attacchi di *hacker* informatici; le possibili derive discriminatorie nei confronti di parte della popolazione senza accesso ad Internet o con un accesso limitato e quelle scaturenti da sempre più pervasive capacità di selezione a seguito di profilazione.

Anche l'impatto sulla competitività del mercato, dopo anni di fiducia indiscriminata, inizia ad essere se non negato, certamente messo in discussione.

Di recente, alcuni studiosi hanno criticato quello che hanno definito come "*welfare mirage*", l'illusione che le nuove tecnologie, con le loro inedite dinamiche di mercato, spingano sempre e comunque verso l'innovazione, la competitività e il beneficio dei consumatori. Secondo questi autori, in realtà, dietro la facciata del mito *naturaliter* pro-competitivo dell'era digitale, occorre analizzare le dinamiche, i rapporti di forza e gli effetti di determinati comportamenti delle tecnologie *Big Data* (algoritmi di *pricing*, profilazione e discriminazione comportamentale, emersione delle *super-platforms* e dei *price comparison website*) ben sapendo che, certo, queste possono avere un effetto pro-competitivo ed apportare benefici ai consumatori, ma tale effetto non è sempre assicurato e/o talvolta il beneficio comporta costi sociali di cui comunque tenere conto.

Insomma, se il mercato è guidato dalla "mano invisibile" che si assume imparziale, la "mano digitalizzata" che guida l'attuale rivoluzione industriale è figlia dell'uomo e, pertanto, tutt'altro che neutra ed esente da manipolazioni.

Tra i diversi fenomeni caratterizzanti l'era digitale uno dei più rilevanti è sicuramento quello costituito dalla possibilità di tradurre in dati digitali fatti, azioni, informazioni, elementi del mondo "reale" (o, per meglio dire, analogico) - "*digital goods*", come li definisce la Commissione Europea. La disponibilità di un numero sterminato di dati - in costante crescita, grazie anche al miglioramento delle tecnologie "digitalizzate" in grado di catturare e creare dati - già da tempo ha conosciuto importanti applicazioni in diversi campi industriali, aprendo scenari che qualche anno fa si pensavano relegati a romanzi fantascientifici, soprattutto grazie all'abbinamento dei Big Data con il modo della robotica, dell'*Artificial Intelligence* (AI) e del c.d. *Internet-of-Things* (IoT).

Se la digitalizzazione sta comportando la progressiva costruzione di un mondo parallelo sempre più completo ed autonomo - un nuovo ecosistema -, il dato digitale appare essere come la forma di vita caratteristica di tale ecosistema.

Una forma di vita difficile da decifrare, rispetto a cui, ad oggi, si è cercato prevalentemente di elencare le caratteristiche più rilevanti piuttosto che offrire definizioni esaustive. Il termine stesso "*Big Data*" è comunemente considerato un iperonimo, una parola più evocativa che definitoria.

Ciò che è certo è che l'unità di misura più adeguata per comprendere la portata del fenomeno non è il singolo dato, ma una quantità massiva caratterizzata dalla velocità di creazione/estrazione, dalla varietà delle fonti e dei contenuti veicolati.

Nel presente elaborato di ricerca si è dato conto di questo processo, per così dire "spontaneo", di costruzione dell'ambiente digitale, per poi concentrarsi sulla principale forma di vita di questo ecosistema, i dati, o meglio, i "Big Data".

Si è fatto riferimento a quanto emerso da ricerche e senza entrare nel merito delle analisi e dei dibattiti - cosa che non ci compete.

Dunque, si è analizzata la questione dell'accesso ai *Big Data* da una prospettiva giuridico-economica.

Come si è potuto notare, i *Big Data* hanno una rilevanza fondamentale negli odierni scenari economici e si è visto che talune barriere limitano l'entrata nei sottomercati dei *Big Data* corrispondenti ai diversi anelli della catena del valore di tali utilità (raccolta,

archiviazione, analisi e uso). Tali barriere all'ingresso costituiscono limiti all'accesso alle risorse digitali, e sono di tipologia differente (tecnologica, giuridica ecc.).

Fra quelle di natura tecnologica, si è visto che soprattutto le esternalità di rete, le economie di scala, di gamma e di velocità, i mercati multi-versante, i costi di transizione (*switching costs*) consentono a uno scarso numero di soggetti economici di raggiungere un notevole vantaggio competitivo determinato dalla concentrazione del potere informativo digitale, che ha effetti negativi sul benessere sociale e sul funzionamento dei vari sottomercati dei *Big Data*. Ulteriori barriere di carattere giuridico, quali quelle relative alla sicurezza dei sistemi di immagazzinamento (che riguardano l'anello dell'archiviazione dei dati) e quelle inerenti a questioni di *privacy* informazionale e protezione dei dati personali (concernenti le fasi della raccolta, dell'analisi e dell'uso dei dati), sono state erette dal legislatore per perseguire determinate finalità sociali, quale la tutela degli interessi degli utenti-consumatori.

Infine, si è passati al vaglio della *privacy* informazionale e della protezione dei dati personali, che riguardano rispettivamente il versante americano e quello europeo.

A differenza del primo, il legislatore europeo ha recentemente stabilito limiti sostanziali all'accesso ai *Big Data* da parte delle imprese e, nel contempo, ha arricchito il novero dei diritti degli interessati, rafforzando, fra gli altri, il diritto di accesso e il diritto alla cancellazione, nonché stabilendo un nuovo diritto alla portabilità dei dati personali. Si è dato conto, inoltre, dei recenti *trends* di commodificazione dei dati personali: benché nessuno dei due ordinamenti presi in esame riconosca un diritto di proprietà su tali utilità, negli Stati Uniti, da un lato, numerosi autori hanno proposto di considerare le informazioni personali al pari di *assets* strategici al fine di regolarizzare il mercato; nell'Unione europea, dall'altro, la protezione dei dati personali non può prescindere dal rango di diritto fondamentale almeno "sulla carta", anche se in alcune norme del Regolamento 679/2016 si ravvisano i germi di una tutela proprietaria.

Le barriere giuridiche all'analisi dei dati personali, riguardanti le attività di utilizzo di algoritmi e sistemi di intelligenza artificiale, costituiscono una questione particolarmente spinosa. Un influente autore statunitense ha proposto di considerare le esternalità negative degli algoritmi alla stregua di immissioni (*nuisances*) di cui gli attori privati che utilizzano le tecniche di analisi dovrebbero farsi carico. Le considerazioni di politica legislativa non possono prescindere, ad avviso di chi scrive,

da un così importante contributo. Nel versante europeo, si dibatte sulle esigenze di tutela dei gruppi di interessati formati sulla base dell'analisi dei dati personali (*group privacy*); inoltre, talune regole del Regolamento (UE) 2016/679 prevedono un diritto alla spiegazione del procedimento decisionale automatizzato.

Un ulteriore limite giuridico all'accesso ai *Big Data* discende dalle questioni di appartenenza inerenti ai dati non personali prodotti dai sensori degli oggetti dell'*Internet delle Cose*, come si è visto (la c.d. *data ownership*). Giacché i regimi esistenti non garantiscono un'adeguata tutela ai *datasets* di ingenti dimensioni, secondo parte della dottrina europea è necessario istituire un nuovo diritto esclusivo su tali *assets*. Nondimeno, come si è ampiamente analizzato nel corso della presente ricerca, una siffatta allocazione giuridica non trova adeguate giustificazioni economiche. Anzi, la previsione di una privativa sui *datasets* finirebbe per rafforzare la posizione di potere (di mercato e politico) di poche piattaforme digitali (quali *Google*, *Amazon*, *Facebook* e *Apple*), basata attualmente sul controllo *de facto* di immensi patrimoni informativi digitali, e pregiudicherebbe l'accesso ai *newcomers* nei sottomercati dei *Big Data*. In questo scenario, occorre promuovere azioni di decentramento delle risorse digitali. È preferibile, dunque, spostare l'attenzione sulla tematica della regolamentazione dell'accesso ai *datasets* mediante altri campi del diritto, quale lo strumentario dell'*antitrust* e la tutela dei consumatori, che però devono "reinventarsi" per raccogliere le sfide imposte dai *Big Data*. Particolarmente ragionevoli paiono le proposte della Commissione di estendere il diritto di portabilità ai dati non personali e prevedere un accesso dietro corrispettivo ai *datasets* mediante la concessione di una licenza obbligatoria a condizioni eque, ragionevoli e non discriminatorie (*Fair, Reasonable And Non-Discriminatory*, FRAND).

Come afferma opportunamente Michal Gal, «*we live in formative times*». Al giurista sarà richiesto un compito difficile in futuro, che vada al di là della mera strategia dello struzzo e consenta di riconquistare gli orizzonti perduti: «*l'innovazione non guarderà indietro*; *sarà necessario capirne i nessi reali, come pure la loro ricaduta sul piano pratico-applicativo*».

Le riforme del quadro giuridico europeo rappresentano una svolta importante per definire un contesto uniforme e proiettato sulle esigenze future e, soprattutto, preservare la fiducia degli utenti nello spazio digitale e nelle sue potenzialità.

Fiducia, innovazione e futuro sono fortemente correlati.

L'obbiettivo al quale dovremmo tendere è la garanzia di uno stesso livello di tutela dei diritti *online* così come *offline*.

Ma se una buona regolamentazione è essenziale, essa non è da sola sufficiente per affrontare l'impatto di questi processi sulle nostre società.

Penso che sia necessaria una nuova consapevolezza da parte delle opinioni pubbliche.

L'attenzione ai *Big Data* non può riguardare soltanto le sue implicazioni scientifiche e tecniche o gli sconvolgenti effetti delle innovazioni sull'economia.

Ci deve preoccupare anche il potenziale discriminatorio che dal loro utilizzo, anche rispetto a dati non identificativi o aggregati, può nascere per effetto di profilazioni sempre più puntuali ed analitiche: in un gioco che finisce per annullare l'unicità della persona, il suo valore, la sua eccezionalità.

E una grande attenzione dobbiamo rivolgere alle applicazioni dell'intelligenza artificiale che effettuano valutazioni o assumono decisioni supportate soltanto da algoritmi, con un intervento umano reso via via più marginale, fino ad annullarsi, con effetti dirompenti sul modo di vivere e articolare esistenze e relazioni, in termini individuali ma anche sociali e politici.

Le Autorità europee di protezione dati avvertono il bisogno di accompagnare questi fenomeni attraverso un più rigoroso approccio etico e di generale responsabilità.

E prima di tutto abbiamo bisogno di promuovere garanzie di trasparenza dei processi, anche per la progressiva difficoltà a mantenere un effettivo controllo sui dati: per l'opacità delle modalità di raccolta, dei luoghi di conservazione, dei criteri di selezione e di analisi.

I rapporti asimmetrici tra chi quei dati fornisce e chi li sfrutta si risolvono a favore di questi ultimi ed in particolare di coloro che gestiscono le piattaforme digitali e dispongono degli standard tecnologici dominanti.

La capacità di elaborare, anche in tempo reale, tramite algoritmi sempre più potenti un'ingente mole di dati consente di estrarre conoscenza e, in misura esponenziale, di effettuare valutazioni predittive sui comportamenti degli individui nonché, più in generale, di assumere decisioni per l'intera collettività.

Chi possiede il profilo dei consumatori indirizza la produzione commerciale verso specifici modelli di utenza, così da assecondarne i gusti ed insieme orientare selettivamente le scelte individuali.

Dobbiamo chiederci quante delle nostre decisioni siano in realtà fortemente condizionate dai risultati che un qualche algoritmo ha selezionato per noi e ci ha messo davanti agli occhi.

Dobbiamo riflettere sugli attuali scenari, interrogarci sugli effetti prodotti da queste trasformazioni: per comprendere le conseguenze sulle nostre vite indotte dalle decisioni automatizzate.

Le tecnologie più progredite incidono ulteriormente sulla fisionomia propria degli Stati, non solo perché ne hanno scardinato il fattore identitario della territorialità, rispetto appunto ad una realtà immateriale che si sviluppa su reti e su sistemi *cloud*, ma perché ne indeboliscono la capacità di conoscere i fenomeni per governarli e di intervenire sulle dinamiche ordinamentali a vantaggio della collettività.

È ovvio che questa profonda crisi non è stata determinata esclusivamente dall'innovazione tecnologica ma è indubbio che tali fenomeni abbiano accelerato queste tendenze.

In parallelo con la crisi dei tradizionali corpi intermedi, le piattaforme digitali, come veri e propri oracoli, paradossalmente, della retorica della disintermediazione sono destinate ad acquistare il ruolo di mediatori della realtà, di interpreti di ciò che accade o di quello che potrà accadere e di assumere decisioni per l'intera collettività.

E nessuno avrà dati in quantità e qualità paragonabili a quelli a loro disposizione mentre in pochi saranno capaci di dare loro un senso utile.

Nel presente lavoro di ricerca si è descritto il punto di vista del giurista e tutto ciò che va sotto il nome di *Data driven innovation* è foriero di nuove esigenze di regolazione che attingono da una realtà altrettanto nuova. L'estrazione di una mole macroscopica di dati stravolge tuttavia le regole del gioco e crea relazioni di potere non evidenti. Tuttavia occorre acquisire la consapevolezza che non possa esservi una teoria giuridica dei *big data* e che l'unica teoria possibile verta proprio sulla impossibilità di una teoria. Un esempio di "teoria dell'assenza" fu il tentativo intellettuale della semantica storica nel secondo dopoguerra in Germania sfociato nella elaborazione dei concetti storici fondamentali, di quei concetti, cioè, che mutavano diacronicamente e non

sincronicamente. Ciò era stato reso possibile attraverso un'acquisita consapevolezza da parte della società circa le proprie trasformazioni ed anche grazie al carattere riflessivo dell'elaborazione concettuale.

In un'epoca in cui parole e cose sono indipendenti, ed anzi, le cose sono incorporate nelle parole, e viceversa, sembra talora anacronistico talora infruttuoso riflettere sui *big data* attraverso le categorie finora utilizzate, in quanto si commetterebbe un errore di prospettiva. Soluzioni come lo *use constraint* con riguardo ai dati estratti dai dispositivi IoT ovvero nuove definizioni di *personal identifiable information* sono potenzialmente efficaci ma necessitano una verifica *ex post.*

Il vantaggio consiste di certo nel fatto che le categorie concettuali, e tra esse le categorie giuridiche, sono esenti - a differenza dei dispositivi tecnologici - dalla cd. "obsolescenza programmata". La loro resistenza al tempo è invece spesso sinonimo di qualità, in quanto essa è indice del carattere omnicomprensivo della tecnica e dell'esperienza giuridica.

Tuttavia l'alcova del giurista è sempre più vulnerabile e le sfide che le tecnologie impongono sono sempre più ardue. Non si può non costatare che le categorie giuridiche siano sempre meno resistenti e che fenomeni come *big data* fagocitino anche i graduali processi di stratificazione della cultura giuridica. A ciò si aggiunga che una delle conseguenze della rapidità con cui flussi così consistenti di dati circolano a livello globale sia la polarizzazione dei meccanismi di tutela, come avviene nel caso della tutela sempre più necessaria e stringente *privacy* dinanzi al dilagare delle tecniche di sorveglianza di massa. Analizzare i fenomeni guardando soltanto agli effetti di medio periodo può risultare inadeguato ai fini dell'interpretazione delle spinte culturali sottostanti. Che si tratti di una nuova cultura sembrerebbe una conclusione prematura, ma certamente *big data* implica lo sviluppo di nuove competenze, sempre più complesse, ed un dialogo serrato tra giuristi e scienziati. Tuttavia il giurista non può sospendere il tempo in attesa della migliore regolazione possibile. Esso deve regolare mentre qualcosa di nuovo accade, prendendo su di sé la responsabilità delle delusioni cognitive.

Nell'elaborazione di modelli di regolazione che non hanno più ad oggetto soltanto leggi ed uomini, bensì "cose intelligenti", appare logico acquisire nuove competenze, che tuttavia non sostituiscano le competenze proprie di un giurista, ma che, al

contrario, invochino l'intervento del giurista nell'ambito di un processo di umanizzazione. Il giurista è dunque chiamato a svolgere, se e quando vi riesce, un ingrato compito, che è quello di acquisire conoscenze tecniche da combinare sapientemente con la cultura giuridica.

Aver in mente che dalle scelte fondamentali di un sistema derivino le soluzioni culturali, implica il dover muoversi sulle sabbie mobili della conoscenza *in fieri*. Per questo è necessario più che mai che il giurista si riappropri del senso dell'importanza dei modelli giuridici e della sua principale competenza, ossia il "fare cose con le parole".

BIBLIOGRAFIA DI RIFERIMENTO

AARON R. ET AL., *Data Brokers In An Open Society*, Open Society Foundations Report, 2016.

ABRAMSON J., *Merchants of Truth*, Simon & Schuster, New York, 2019.

ABRIANI N. ET. AL., *Diritto industriale*, in *Trattato di Diritto Commerciale*, diretto da COTTIMO G., Cedam, 2001.

ACCENTURE STRATEGY, *Reworking the Revolution: Are you ready to compete as intelligent technology meets human ingenuity to create the future workforce?*, 2018.

AMATO G., *Il potere e l'antitrust*, Il Mulino, 1998.

ANDERSON C., *The End of Theory: the Data Deluge Makes the Scientific Method Obsolete*, «Wired», 2008.

ANDREJEVIC M., *iSpy: Surveillance and Power in the Interactive Era*, University Press of Kansas, 2007.

ANGIULI O. ET. AL., *How to De-Identify Your Data. Balancing statistical accuracy and subject privacy in large social-science data sets*, in *ACM Queue*, 2015.

APLIN T., *A Critical Evaluation of the Proposed EU Trade Secrets Directive*, King's College London, Dickson Poon School of Law Legal Studies, Research Paper n. 25, 2014.

ARCIDIACONO D., *The Trade Secrets Directive in the International Legal Framework*, in *European Papers*, 2016.

ARICKER M. - MCGUIRE T. - PERRY J., *Harvard Business Review: Five Roles You Need on Your Big Data Team*, New York, 2013.

ARMSTRONG M., *Competition in Two-Sided Markets*, in *Rand Journal Of Economics*, 2006.

ARVIDSSON A. - BONINI T., *Valuing Audience Passions: From Smythe to Tarde*, in *European Journal of Cultural Studies*, New York, 2015.

AUTERI P. ET AL., *Diritto industriale: proprietà intellettuale e concorrenza*, Giappichelli, Torino, 2016.

BAGNOLI V., *The Big Data relevant market*, in *Concorrenza e Mercato*, n. 23, 2016.

BALGANESH S., *Quasi-property: like, but not quite property*, in *University of Pennsylvania Law Review*, 2012.

BALKIN J.M., *The Three Laws of Robotics in the Age of Big Data*, in *Ohio State Law Journal*, 2017.

BANTERLE F., *Brevi cenni sulla titolarità dei dati comportamentali nei Big Data tra privacy e proprietà intellettuale*, in *AIDA*, 2016.

BARBARO M. – ZELLER T., 1 *A Face Is Exposed for AOL Searcher No. 4417749*, in The New York Times, 9 agosto 2016.

BARNEY D. ET AL., *The Participatory Condition in the digital age*, University of Minnesota Press, 2016.

BARRETT N., *Will Big Data create a new untouchable business elite?*, in *The Economist*, 2017.

BAUMAN Z. - LYON D., *Liquid Surveillance: A Conversation*, Polity Press, 2012.

BAYLES M.E., *Principles of Law: A Normative Analysis*, Springer, 2016.

BENKLER Y., *Degrees of Freedom, Dimensions of Power*, in *Daedalus*, 2016.

BERGELSON V., *It's Personal But Is It Mine? Toward Property Rights in Personal Information*, in *University of California Davis Law Review*, 2013.

BERK R., *The role of race in forecast of violent crime*, in *Race and Social Problems*, 2018.

BERNERS-LEE T., *L'architettura del nuovo web: dall'inventore della rete il progetto di una comunicazione democratica, interattiva e intercreativa*, Feltrinelli, Milano, 2018.

BESANKO D. - BRAEUTIGAM R.R., *Microeconomics*, Wiley & Sons, 2014.

BEVYERS E. ET AL., *Räume und Kulturen des Privaten*, Springer, 2017.

BINNS R., *Data protection impact assessments: a meta-regulatory approach*, in *International Data Privacy Law*, 2017.

BLAIR R.D. - SOKOL D., *Cambridge Handbook of Antitrust, Intellectual Property and High Tech*, Cambridge University Press, 2018.

BORGHI M. - KARAPAPA S., *Contractual restrictions on lawful use of information: databases protected by the back door?*, in *European Intellectual Property Review*, 2015.

BORK R., *The Antitrust Paradox*, Free Press, 2016.

BOSTON CONSULTING GROUP, *Global Retail Banking 2018: The Power of Personalization*, 2018.

BOURREAU M. - DE STREEL A. - GRAEF I., *Big Data and competition policy: market power, personalised pricing and advertising*, in *Cerre project report*, 2017.

BOYD D. - CRAWFORD K., *Critical Questions for Big Data: Provocations for a cultural, technological, and scholarly phenomenon*, in *Information, Communication & Society*, 2012.

BRAITHWAITE J. - DRAHOS P., *Global Business Regulation*, Cambridge University Press, 2017.

BRANDIMARTE L. ET AL., *Misplaced Confidences: Privacy and the Control Paradox*, in *Social Psychological and Personality Science*, 2012.

BROTHERTON R. - GIACONE G. L., *Menti sospettose. Perché siamo tutti complottisti*, Bollati Boringhieri, Torino, 2017.

BRYNJOLFSSON E. - MITT L. - KIM H., *Strength in Numbers: How does Data-driven Decision Making Affect Firm Performance?*, in *Social science research network paper*, 2011.

BRYNJOLFSSON E. - MITT L. - KIM H., *Strength in Numbers: How does Data-driven Decision Making Affect Firm Performance?*, in *Social science research network paper*, 2015.

BURRELL J., *How the machine 'thinks': Understanding opacity in machine learning algorithms*, in *Big Data & Society*, 2016.

CALABRESI G., *Property Rules, Liability Rules, and Inalienability: One View of the Cathedral*, Harvard Law Review, Harvard, 2010.

CASTELLS M., *The Information Age: Economy, Society and Culture. Volume I: The Rise of the Network Society*, Wiley Blackwell, 2010.

CATE F.H. - MAYER-SCHÖNBERGER V., *Notice and consent in a world of Big Data*, in *International Data Privacy Law*, 2013.

CATE F.H., *The failure of fair information pratice principles*, in WINN J.K. (a cura di), *Consumer protection in the Age of the «Information Economy»*, Ashgate, Aldershot, 2016.

CAVANILLAS J.M. - CURRY E. - WAHLSTER W., *New Horizons for a Data-Driven Economy. A roadmap for Usage and Exploitation of Big Data in Europe*, Springer, 2016.

CHANDLER A.D., *Scale and Scope. The Dynamics of Industrial Capitalism*, Harvard University Press, 2010.

CHANDLER A.D., *The Visible Hand. The Managerial Revolution in American Business*, Harvard University Press, 2014.

CHEN M. ET AL., *Big Data: Related Technologies, Challenges and Future Prospects*, Springer, 2014.

CHRISTIAN B. - GRIFFITHS T., *Algorithms to live by*, William Collins, 2016.

CIANI J., *Property rights model v. contractual approach: how protecting non personal data in cyberspace?*, in *Dir. Comm. Int*., 2017.

COASE R.H., *The problem of social cost*, in *Journal of Law & Economics*, 2013.

CODD E.F., *A Relational Model of Data for Large Shared Data Banks*, Addison Wesley, 2014.

COHEN J., *Examined Lives: Informational Privacy and the Subject as Object*, in *Stanford Law Review*, 2010.

COHEN J., *What Privacy is For*, in *Harvard Law Review*, 2013.

COLANGELO G. - FALCE V., *Concorrenza e comportamenti escludenti nei mercati dell'innovazione*, Il Mulino, Bologna, 2017.

COLANGELO G. - MAGGIOLINO M., *Big Data as a misleading facility*, in *European Competition Journal*, 2017.

COLANGELO G., *Big Data, piattaforme digitali e antitrust*, in *Mercato concorrenza regole*, n. 3, 2016.

COOTER R. - ULENT T., *Law & Economics*, Addison-Wesley, 2012.

CRAIN M., *The limits of transparency: Data brokers and commodification*, City University of New York, Academic Works, 2017.

DAVENPORT T.H. – BECK J.C., *The Attention Economy: Understanding The New Currency Of Business*, Harvard Business School Press, 2011.

DAVIS K. - PATTERSON D., *Ethics of Big Data*, O'Reilly, 2012.

DAVISON M., *The Legal Protection of Databases*, Cambridge University Press, 2013.

DE FRANCESCHI A. - LEHMANN M., *Data as Tradeable Commodity and New Measures for their Protection*, in *Italian Law Journal*, 2015.

DE HERT P. - PPAPKOSTANTINOU V., *The proposed data protection Regulation replacing Directive 95/46/EC: A sound system for the protection of individuals*, in *Computer Law & Security Review*, 2012.

DE MAURO A. ET AL., *A formal definition of Big Data based on its essential features*, in *Library Review*, 2016.

DE MAURO A., *Big Data Analytics. Analizzare e interpretare dati con il machine learning*, Apogeo, Milano, 2019.

DE MONTJOYE Y. ET AL., *Unique in the Crowd: The privacy bounds of human mobility*, in *Scientific Reports*, 2013.

DERCLAYE E., *The Legal Protection of Databases: A Comparative Analysis*, Edward Elgar, 2008.

DI PORTO F., *La regolazione degli obblighi informativi*, Editoriale Scientifica, Napoli, 2017.

DIGITAL CULTURE MEDIA AND SPORT COMMITTEE, *Disinformation and «FakeNews»: Final Report*, London, 18 febbraio 2019.

DOURISH P., *Algorithms and their others: Algorithmic culture in context*, in *Big Data & Society*, 2016.

DRAHOS P. - BRAITHWAITE J., *Information Feudalism: Who Owns the Knowledge Economy?*, Earthscan Publications Ltd, 2012.

DREIER T., *Online and Its Effect on the "Goods" Versus "Services" Distinction*, in *International Review of Intellectual Property and Competition Law*, 2013.

DREXL J., *Designing competitive markets for industrial data. Between Propertisation and Access*, Max Planck Institute for Innovation and Competition, Research Paper nn. 16-13, 2016.

DREYFUSS R.C. - STRANDBURG K.J., *The Law And Theory Of Trade Secrecy: A Handbook of Contemporary Research*, Edward Elgar, 2011.

DUGAIN M. - LABBÉ C., *L'uomo nudo. La dittatura invisibile del digitale*, Enrico Damiani Editore, 2016.

DUMBILL E., *Making sense of Big Data*, in *Big Data*, 2013.

DURANTE M. - PAGALLO U., *Manuale di informatica giuridica e diritto delle nuove tecnologie*, Utet Giuridica, Torino, 2012.

DURHAM M.G. - KELLNER D.M., *Media and Cultural Studies: Keyworks*, Wiley-Blackwell, 2012.

ERNST & YOUNG, *As FinTech becomes the norm, you need to stand out from the crowd*, Global FinTech Adoption Index, 2019.

ESTEVE A., *The business of personal data: Google, Facebook, and privacy issues in the EU and the USA*, in *International Data Privacy Law*, 2017.

EVANS D. - SCHMALENSEE R., *Markets with Two-Sided Platforms*, in *Issues In Competition Law And Policy*, 2014.

EVANS D., *Attention to Rivalry among Online Platforms and Its Implications for Antitrust Analysis*, Coase-Sandor Institute for Law & Economics, Working Paper n. 627, 2013.

EVANS D. - SCHMALENSEE R., *Matchmakers: The New Economics of Multisided Platforms*, Harvard Business School Pr, 24 maggio 2016.

EZRACHI A. - STUKE M.E., *Artificial intelligence and collusion: when computers inhibit competition*, in *Working paper CCLP*, 2015.

FALCE V., *Trade Secret Protection in the Innovation Union. From the Italian approach to the UE solution*, in *Mercato, concorrenza e regole*, 2013.

FARKAS T.J., *Data Created by the Internet of Things: The New Gold without Ownership*, in *Revista La Propiedad Inmaterial*, 2017.

FERTIK M. - THOMPSON D., *The Reputation Economy. How To Optimise Your Digital Footprint In a World Where Your Reputation Is Your Most Valuable Asset*, Crown Business, 2015.

FLEISH E., *What is the Internet of things? An economic perspective*, in *Economics, management and financial markets*, 2010.

FLORIDI L., *Big Data and Their Epistemological Challenge*, in *Philosophy and Technology*, 2012.

FLORIDI L., *La quarta rivoluzione*, Raffaello Cortina, Milano, 2014.

FLORIDI L., *Open data, data protection, and group privacy*, in *Philosophy and Technology*, 2014.

FLORIDI L., *Protection of Information and the Right to Privacy. A New Equilibrium?*, Springer, 2014.

FLORIDI L., *The Ethics of Information*, Oxford University Press, 2013.

FLORIDI L., *The Fourth Revolution: How the Infosphere is Reshaping Human Re-ality*, Oxford University Press, 2014.

FONTANA F. - CAROLI M., *Economia e Gestione delle Imprese*, McGraw-Hill, Milano, 2013.

FORAY D., *L'economia della conoscenza*, Il Mulino, 2006.

FORD M., *Industry 4.0: Making the first move*, SMT magazine, 2016.

FREEMAN J., *The Private Role in Public Governance*, in *New York Law Review*, 2010.

FRIEDMAN D. ET AL., *Some Economics of Trade Secret Law*, in *The Journal of Economic Perspectives*, 2010.

FRIEDMAN L.M., *Guarding Life's Dark Secrets: Legal and Social Controls over Reputation, Propriety, and Privacy*, Stanford University Press, 2011.

FRIEDMAN L.M., *Name Robbers: Privacy, Blackmail, and Assorted Matters in Legal History*, in *Hofstra Law Review*, 2012.

FUGGETTA A., *Cittadini ai tempi di internet*, Franco Angeli, Milano, 2018.

OLIVIERI G., *Informazione e Big Data tra innovazione e concorrenza*, Giuffrè Editore, Milano, 2018.

GAMBARO A., *Dai beni immobili ai beni virtuali*, in *Enciclopedia Treccani*, 2017.

GAMBARO A., *Trattato dei diritti reali. Volume 1: proprietà e possesso*, Giuffrè, Milano, 2011.

GARANTE PER LA PROTEZIONE DEI DATI PERSONALI, *Big Data: Agcom, Antitrust e Garante privacy avviano indagine conoscitiva*, 2017.

GAVAZZI G., *Norme primarie e norme secondarie*, Giappichelli, 2011.

GELLMAN R., *Fair Information Practices: A Basic History*, in *SSRN Library*, 2017.

GILBERT MILLER H. - MORK P., *From Data to Decisions: A Value Chain for Big Data*, in *IT Professional*, 2013.

GITELMAN L., *"Raw Data" is an Oxymoron*, MIT Press, 2013.

GOODMAN B. - FLAXMAN S., *European Union regulations on algorithmic decision-making and a "right to explanation"*, in *AI Magazine*, 2017.

GOODMAN B., *A Step Towards Accountable Algorithms?: Algorithmic Discrimination and the European Union General Data Protection*, 29th Conference on Neural Information Processing Systems Paper, 2016.

GORZ A., *L'immateriale: Conoscenza, Valore e Capitale*, Bollati Boringhieri, 2013.

GOSNEY M., *5 ways industry 4.0 could change your factory*, 2015.

GRAEF I., *Market definition and market power in data: the case of online platforms*, in *World Competition Law and Economics*, 2015.

GRANT R.M., *Toward a knowledge-based theory of the firm*, in *Strategic Management Journal*, 2018.

GRECO P., *I diritti sui beni immateriali*, Utet Giuridica, 2013.

GREENBERGER M., *Computers, Communications, And The Public Interest*, John Hopkins Press, 2017.

GRUNES A. - STUCKE M.E., *No Mistake About It: The Important Role of Antitrust in the Era of Big Data*, in *The Antitrust Source*, 2015.

GUEST A.G., *Oxford Essays in Jurisprudence*, Oxford University Press, 2012.

GURIN J., *Big Data and Open Data: How Open Will the Future Be?*, in *Journal of Law and Policy for the Information Society*, 2015.

HARARI Y.N., *21 lezioni per il XXI secolo*, Bompiani, Milano, 2018.

HARDIN G., *The Tragedy of the Commons*, in *Science*, 2014.

HARFORD T., *Big data are we making a big mistake?*, in *Financial Times*, 28 marzo 2014.

HARTMANN P. - ZAKI M. - FIELDMANN N. - NEELY A., *University of Cambridge: Big Business? A taxonomy of Data-driven Business Models used by Start-up firms*, marzo 2014.

HEGEL G.W.F., *Lineamenti di filosofia del diritto*, Bompiani, 2016.

HELLER M., *The Tragedy of the Anticommons: Property in the Transition from Marx to Markets*, in *Harvard Law Review*, 2016.

HESS C. - OSTROM E., *Understanding Knowledge as a Commons*, MIT Press, 2017.

HILDEBRANDT M. - DE VRIES K., *Privacy, Due Process and the Computational Turn-Philosophers of Law Meet Philosophers of Technology*, Routledge, 2013.

HILL R.K., *What an Algorithm Is*, in *Philosophy & Technology*, 2016.

HOOFNAGLE C. - WHITTINGTON J., *Free Accounting for the Costs of the Internet's Most Popular Price*, in *UCLA Law Review*, 2014.

HOSNI H. - VULPIANI A., *Forecasting in Light of Big Data*, in *Philosophy & Technology*, 2017.

HOWARD P., *Pax Technica. How the Internet of Things May Set Us Free or Lock Us Up*, Yale University Press, 2015.

HULL G., *Digital Copyright and the Possibility of Pure Law*, in *Qui Parle*, 2013.

JAMES W., *The Principles of Psychology*, Henry Holt and Company, 2011.

KAYE L., *The proposed EU Directive for the legal protection of databases: a cornerstone of the information society?*, in *European Intellectual Property Review*, 2008.

KERBER W., *A New (Intellectual) Property Right for Non-Personal Data? An Economic Analysis*, Macie Discussion Paper n. 3-2016, 2016.

KERBER W., *Exhaustion of Digital Goods: An Economic Perspective*, Macie Discussion Paper n. 23, 2016.

KERBER W., *Governance of Data: Exclusive Property vs. Access*, in *IIC International Review of Intellectual Property and Competition Law*, 2016.

KIRKPATRICK D., *The Facebook effect*, Simon & Schuster, 2015.

KLIAZOVICH D. ET AL., *GreenCloud: a packet-level simulator of energy-aware cloud computing data centers*, in *Journal of Supercomputing*, 2014.

KLIESEN K.L. - MCCRACKEN M.W., *Tracking the U.S. Economy with Nowcasts*, in *The Regional Economist*, 2016.

KLIMAS T. - VAICIUKAITE J., *The Law of Recitals in European Community Legislation*, in *ILSA Journal of International & Comparative Law*, 2008.

KOOMEY J.K., *Worldwide electricity used in data centers*, in *Environmental Research Letters*, 2013.

KOOPS B.J., *Forgetting footprints, shunning shadows. A critical analysis of the "right to be forgotten" in Big Data practice*, in *SCRIPTed*, 2013.

KOSINSKI M. ET AL., *Private traits and attributes are predictable from digital records of human behavior*, in *PNAS*, 2013.

KROLL J.A. ET AL., *Accountable Algorithms*, in *University of Pennsylvania Law Review*, 2017.

KUEMPEL A., *The Invisible Middleman: A Critique and Call for Reform of the Data Broker Industry*, in *Northwestern Journal of International Law & Business*, 2016.

LANDES W. - POSNER R., *The Economic Structure of Intellectual Property Law*, Harvard University Press, 2013.

LANEY D., *3D Data Management*: *Controlling Data Volume, Velocity, and Variety*, Meta Group, 2011.

LEENES R. ET AL., *Data Protection and Privacy: (In)visibilities and Infrastructures*, Springer, 2017.

LEMLEY M., *IP In a World Without Scarcity*, in *New York University Law Review*, 2015.

LEMLEY M., *Private Property*, in *Stanford Law Review*, 2014.

LEMLEY M., *Property, Intellectual Property, and Free Riding*, John M. Olin Program in *Law and Economics Working Paper*, n. 291, 2014.

LEMLEY M., *The Surprising Virtues of Treating Trade Secrets as IP rights*, in *Stanford Law Review*, 2011.

LESSING L., *Code and Other Laws of the Cyberspace*, Basic Books, 2016.

LESSIG L., *Cultura libera. Un equilibrio fra anarchia e controllo, contro l'estremismo della proprietà intellettuale*, Apogeo, Milano, 2015.

LESSIG L., *Privacy as Property*, in *Social Research*, 2012.

LESSIG L., *The Architecture of Privacy*, in *Vanderbilt Entertainment Law and Practice*, 2016.

LÉVEQUE F. - MÉNIÈRE Y., *The Economics of Patents and Copyright*, Berkeley University Press, 2014.

LÉVEQUE F. - SHELANSKI H., *Antitrust, Patents and Copyright: EU and US Perspectives*, Edward Elgar, 2015.

LIBERTINI M., *Le informazioni aziendali riservate (segreti commerciali) come oggetto di diritti di proprietà industriale*, in *Dir. ind*., 2017.

LIPPMANN W., *L'opinione pubblica*, Donzelli Editore, Roma, 2015.

LITMAN J., *Information Privacy/Information Property*, in *Stanford Law Review*, 2010.

LOBEL O., *The Renew Deal: The Fall of Regulation and the Rise of Governance in Contemporary Legal Thought*, in *Minnesota Law Review*, 2014.

LOOS M., *The Regulation of Digital Content B2C Contracts in CESL*, Centre for the Study of European Contract Law, Working Paper Series n. 10, 2013.

LUGARD P. - ROACH L., *The era of Big Data and EU/U.S. divergence for refusal to deal*, in *Antitrust*, Vol. 31, 2017.

LUNDQVIST B., *"Turning Government Data Into Gold": The Interface Between EU Competition Law and the Public Sector Information Directive-With Some Comments on the Compass Case*, in *IIC International Review of Intellectual Property and Competition Law*, 2015.

LUNDQVIST B., *Big Data, Open Data, Privacy Regulations, Intellectual Property and Competition Law in an Internet of Things World. The Issue of Accessing Data*, Faculty of Law, University of Stockholm, Research Paper n. 1, 2016.

LYCETT M., *"Datafication": making sense of (big) data in a complex world*, in *European Journal of Information Systems*, 2018.

LYON D., *Surveillance Studies: An Overview*, Polity Press, 2007.

LYON D., *Surveillance, Snowden, and Big Data: Capacities, consequences, critique*, in *Big Data and Society*, 2014.

MACNISH K.N.J., *Unblinking Eyes: The Ethics of Automated Surveillance*, in *Ethics and Information Technology*, 2016.

MAGGIOLINO M., *Big Data e prezzi personalizzati*, in *Concorrenza e mercato*, 2016.

MALGIERI G., *Trade Secrets v Personal Data: a possible solution for balancing rights*, in *International Data Privacy Law*, 2016.

MANTELERO A., Big Data*: i rischi della concentrazione del potere informativo digitale e gli strumenti di controllo*, in *Diritto dell'informazione e dell'informatica*, 2012.

MANTELERO A., *Personal data for decisional purposes in the age of analytics: From an individual to a collective dimension of data protection*, in *Computer Law & Security Review*, 2016.

MANYIKA J. - CHUI M. - BROWN B., *Big Data: The next frontier for innovation, competition, and productivity*, Mckinsey Global Institute, 2011.

MARR B., *Managing and Delivering Performance*, Routledge, 2016.

MATTEI U., *Il modello di Common Law*, Giappichelli, 2014.

MATTIOLO M., *Disclosing Big Data*, in *Minnesota Law Review*, 2014.

MAYER-SCHÖNBERGER V. - CUKIER K., *Big data. Una rivoluzione che trasformerà il nostro modo di vivere e già minaccia la nostra libertà*, Garzanti, Milano, 2013.

MAYER-SCHÖNBERGER V. - PADOVA Y., *Regime Change? Enabling Big Data through Europe's new Data Protection Regulation,* in *Columbia Science & Technology Law Review,* 2016.

MAYER-SCHÖNBERGER V., *Delete: The virtue of forgetting in the digital age*, Princeton University Press, 2017.

MAZZIOTTI G., *EU Digital Copyright Law and the End-User,* Springer, 2016.

MCDONALD A.M. - CRANOR L.F., *The Cost of Reading Privacy Policies*, in *I/S: A Journal of Law and Policy for the Information Society*, 2015.

MCGUIGAN L. - MANZEROLLE V., *The Audience Commodity in a Digital Age: Revisiting a Critical Theory of Commercial Media,* Peter Lang, 2014.

MEHRA S.K., *Competition Law for a Post-Scarcity World*, in *Texas A&M Law Review*, 2016.

MÉNIÈRE Y. - THUMM N., *Fair, Reasonable and Non-Discriminatory (FRAND) Licensing Terms. Research Analysis of a Controversial Concept*, JRC Science and Policy Report, 2015.

MICKLITZ N.W. - REICH N., *The Commission Proposal for a 'Regulation on a Common European Sales Law (CESL)' - Too Broad or Not Broad Enough?*, EUI Working Papers LAW n. 4, 2012.

MIGLIETTI L., *Profili storico-comparativi del diritto alla privacy*, in *Rivista di diritti comparati,* 2017.

MITTELSTADT B.D. ET AL., *The ethics of algorithms: Mapping the debate,* in *Big Data & Society*, 2016.

MITTELSTADT B.D., *Auditing for Transparency in Content Personalization Systems*, in *International Journal of Communication,* 2016.

MOLINO J.L. - SEDKAOUI S., *Big Data, Open Data and Data Development*, ISTE Ltd. and Wiley & Sons, 2016.

MOROZOV E., *Internet non salverà il mondo. Perché non dobbiamo credere a chi pensa che la Rete possa risolvere ogni problema*, Mondadori, Milano, 2014.

MUNZER S., *New Essays in the Legal and Political Theory of Property*, Cambridge Studies in Philosophy and Law, 2011.

MUNZER S., *A Theory of Property*, Cambridge University Press, 2016.

MURPHY R.S., *Property Rights in Personal Information: An Economic Defence of Privacy,* in *Georgetown Law Journal*, 1995.

MYSKA M. - HARASTA J., *Less is more? Protecting Databases in the EU After Ryanair*, in *Masaryk University Journal of Law and Technology*, 2016.

NELSON R.R., *The Rate and Direction of Inventive Activity: Economic and Social Factors*, Princeton University Press, 2015.

NEWMAN J.M., *Antitrust in zero-price markets: Foundations*, in *University of Pennsylvania Law Review*, 2015.

NEWMAN N., *Search, Antitrust and the Economics of the Control of User Data*, in *Yale Journal on Regulation*, 2014.

NICITA A. ET AL., *Le opzioni nel mercato delle regole*, SIDE Working Paper, 2015.

NORDHAUS W.D., *Productivity growth and the new economy*, Brookings Papers on Economic Activity, 2012.

NOTO LA DIEGA G., *Software Patents and the Internet of Things in Europe, the United States and India*, in *European Intellectual Property Review*, 2017.

O'BRIEN D., *The Right of Privacy*, in *Columbia Law Review*, 2010.

O'LEARY D., *Big Data, Internet of Things and the Internet of Signs*, in *Intelligent Systems in accounting, finance and management*, 2013.

O'NEILL C., *Armi di distruzione matematica, Come i Big Data aumentano la disuguaglianza e minacciano la democrazia*, Bompiani, Milano, 2016.

OHLHORST F., *Big Data Analytics. Turning Big Data Into Big Money*, John Wiley & Sons, 2013.

OHM P., *Broken promises of privacy: responding to the surprising failure of anonymization*, in *UCLA Law Review*, 2017.

OREFICE M., *I Big Data. Regole e concorrenza*, in *Politica del diritto*, 2016.

OSTROM E., *Governing the Commons. The Evolution of Institutions for Collective Action*, Cambridge University Press, 2013.

OTTOLIA A., *Big Data e innovazione computazionale*, Utet, Torino, 2017.

OVI A. - JACOBELLI G.P., *Lo sfruttamento dei dati digitali*, in *MIT Technology Review*, 2013.

PAEZ M. - LA MARCA M., *The Internet Of Things: Emerging Legal Issues For Businesses*, in *Northern Kentucky Law Review*, 2016.

PAGALLO U., *Algo-Rhythms and the Beat of the Legal Drum*, in *Philosophy & Technology*, 2017.

PAGALLO U., *Il diritto nell'età dell'informazione. Il riposizionamento tecnologico degli ordinamenti giuridici tra complessità sociale, lotta per il potere e tutela dei diritti*, Giappichelli, 2014.

PAGALLO U., *The Legal Challenges of Big Data: Putting Secondary Rules First in the Field of EU Data Protection*, in *European Data Protection Law Review*, 2017.

PARRISH A., *The Effects Test: Extraterritoriality's Fifth Business*, in *Vanderbilt Law Review*, 2008.

PASQUALE F., *The Black Box Society. The Secret Algorithms That Control Money and Information*, Harvard University Press, 2017.

PAVOLOTSKY J., *Privacy in the age of Big Data*, in *The Business Lawyer*, 2013.

PENNER J.E., *The "Bundle of Rights" Picture of Property*, in *UCLA Law Review*, 2012.

PERSSON P., *Attention Manipulation and Information Overload: Barriers to Consumer Protection*, Behavioral Public Policy, 2018.

PERZANOWSKY A. - SCHULTZ J., *Legislating Digital Exhaustion*, in *Berkeley Technology Law Journal*, 2014.

PITRUZZELLA G., *Big Data, Competition and Privacy: A Look from the Antitrust Perspective*, in *Concorrenza e mercato*, 2016.

PORTER M.E., *Competitive Advantage: creating and sustaining superior Performance*, Free Press, 1985.

PROSSER W.L., *Handbook on the Law of Torts*, West Pub. Co., 2014.

PROSSER W.L., *Privacy*, in *California Law Review*, 2011.

PURTOVA N.N., *Property rights in personal data: A European perspective*, BOXPress BV, 2011.

PURTOVA N.N., *Property rights in personal data: Learning from the American discourse*, in *Computer Law and Security Review*, 2017.

QUATTROCIOCCHI W. - VICINI A., *Misinfarmation*, Franco Angeli, Milano, 2016.

QUERY T., *Grade infation and the Good-Student Discount*, in *Contingencies Magazine*, American Academy of Actuaries, maggio-giugno 2017.

QUINTARELLI S., *Capitalismo immateriale. Le tecnologie digitali e il nuovo conflitto sociale*, Bollati Boringhieri, Torino, 2019.

RADDEN KEEFE P., *Can network theory thwart terrorist?*, in *The New York Times*, 12 marzo 2016.

RADIN M.J., *Property and Personhood*, in *Stanford Law Review*, 2012.

RATLIFF J. - RUBINFELD D., *Is there a market for organic search engine results and can their manipulation give rise to antitrust liability?*, in *Journal of Competition Law and Economics*, 2014.

RAYPORT J.F. - SVIOKLA J.J., *Exploiting the virtual value chain*, in *Harvard Business Review*, 2017.

REICHMAN J.H. - SAMUELSON P., *Intellectual Property Rights in Data?*, in *Vanderbilt Law Review*, 2016.

RESTA G., *Diritti esclusivi e nuovi beni immateriali*, Utet Giuridica, 2011.

REZZANI A., *Big data. Architettura, tecnologie e metodi per l'utilizzo di grandi basi di dati*, Apogeo, Milano, 2013.

RICH M.L., *Machine Learning, Automated Suspicion Algorithms, And The Fourth Amendment*, in *University of Pennsylvania Law Review*, 2016.

RICHARDS N.M. - KING J., *Big Data Ethics*, in *Wake Forest Law Review*, 2014.

RICHARDS N.M. - SOLOVE D.J., *Prosser's Privacy Law: A Mixed Legacy*, in *California Law Review*, 2010.

RICHARDS N.M., *Intellectual Privacy*, in *Texas Law Review*, 2008.

RICHARDS N.M., *The Dangers of Surveillance*, in *Harvard Law Review*, 2013.

RICHARDS N.M., *The Limits of Tort Privacy*, in *Journal of Telecommunications and High Technology Law*, 2011.

RIFKIN J., *L'era dell'accesso, La rivoluzione della new economy*, Mondadori, Milano, 2010.

RIFKIN J., *Società a costo marginale zero*, Mondadori, 2014.

ROCCASALVA G., *I Big Data e gli strumenti di visualizzazione analitica: interazioni e studi induttivi per le P.A.*, Apogeo, Milano, 2018.

ROCHET J.C. – TIROLE J., *Platform Competition in Two-Sided Markets*, in *Journal of European Economic Association*, 2003.

RODOTÀ S., *Intervista su privacy e libertà*, Laterza, Bari, 2005.

ROSE C.M., *Romans, Roads, and Romantic Creators: Traditions of Public Property in the Information Age*, in *Law & Contemporary Problems*, 2013.

ROSEN J., *The right to be forgotten*, in *Stanford Law Review Online*, 2012.

RUBINFELD D.L. - GAL M.S., *Access Barriers to Big Data*, in *Arizona Law Review*, 2017.

RUBISTEIN I., *Big Data: The End of Privacy or a New Beginning?*, in *International Data Privacy Law*, 2013.

RULLANI E., *Le capitalisme cognitif: du déjà vu?*, in *Multitudes*, 2017.

RYSMAN M., *The Economics of Two-sided Markets*, in *Journal Of Economic Perspectives*, 2017.

SACCO R., *Introduzione al diritto comparato*, Utet Giuridica, Torino, 2016.

SAMUELSON P., *Privacy as Intellectual Property*, in *Stanford Law Review*, 2014.

SARTOR G., *The right to be forgotten in the Draft Data Protection Regulation*, in *International Data Privacy Law*, 2015.

SARTOR G., *The right to be forgotten: balancing interests in the flux of time*, in *International Journal of Law and Information Technology*, 2016.

SARTORE F., *Big Data: Privacy and Intellectual Property in a Comparative Perspective*, Trento Law and Technology, Research Group Student Paper n. 26, 2016.

SCHEPP N.P. - WAMBACH A., *On Big Data and Its Relevance for Market Power Assessment*, in *Journal of European Competition & Practice*, 2016.

SCHOLZ L., *Privacy as Quasi-Property*, in *Iowa Law Review*, 2016.

SCHWARTZ P.M. - JANGER E.J., *Notification of data security breaches*, in *Michigan Law Review*, 2017.

SCHWARTZ P.M. - SOLOVE D., *Reworking Information Privacy Law: A Memorandum Regarding Future ALI Projects About Information Privacy Law*, 2012.

SCHWARTZ P.M., *Property, Privacy, and Personal Data*, in *Harvard Law Review*, 2014.

SCHWARTZ P.M., *The EU-U.S. privacy collision: A turn to institutions and procedures*, in *Harvard Law Review*, 2013.

SCOTT G., *The Four. I padroni: Il DNA segreto di Amazon, Apple, Facebook e Google*, Hoepli, Milano, 2018.

SHELANSKI H., *Information, innovation, and competition policy for the Internet*, in *University of Pennsylvania Law Review*, 2013.

SHORT J.L., *The Paranoid Style in Regulatory Reform*, in *Hastings Law Journal*, 2012.

SIMON P., *Wired: Big Data Lessons from Netflix*, 2014.

SINGEL R., *Netflix spilled your brokeback mountain secret*, Lawsuite Claims, in *Wired*, 17 dicembre 2016.

SLOMAN S. - FERNBACH P., *The Knowledge Illusion: Why We Never Think Alone*, Riverhead Books, 2018.

SMITH E.E. - KOSSLYN S.M., *Cognitive Psychology: Mind And Brain*, Pearson, 2017.

SMYTHE D.W., *Communications: Blindspot of Western Marxism*, in *CTheory*, 2015.

SOLOVE D.J. - RICHARDS N.M., *Privacy's Other Path: Recovering the Law of Confidentiality*, in *Georgetown Law Journal*, 2012.

SOLOVE D.J. - SCHWARTZ P.M., *Information Privacy Law*, Wolters Kluwer, 2015.

SOLOVE D.J., *A Taxonomy of Privacy*, in *University of Pennsylvania Law Review*, 2016.

SOLOVE D.J., *Introduction: Privacy Self-Management and the Consent Dilemma*, in *Harvard Law Review*, 2013.

SOLOVE D.J., *Understanding Privacy*, Harvard University Press, 2008.

SORO A., *Persone in rete*, Fazi, Roma, 2018.

SPADA P., *Conclusioni al Convegno su "IP e Costituzioni" organizzato presso l'Università di Pavia il 23 e 24 settembre 2005*, in *AIDA*, 2015.

STAMATOUDI I., *The EU Databases Directive: Reconceptualising Copyright and Tracing the Future of the Sui Generis Right*, in *Revue hellénique de droit international*, 2015.

STRAHILEVITZ L., *Toward a Positive Theory of Privacy Law*, in *Harvard Law Review*, 2013.

STUCKE M.E. - GRUNES A.P., *Big Data and Competition Policy*, Oxford University Press, 2016.

SUNSTEIN C.R., *Conformity: The Power of Social Influences*, NYU Press, 2019.

SURBLYTÉ G., *Data Mobility at the Intersection of Data, Trade Secret Protection and the Mobility of Employees in the Digital Economy*, Max Planck Institute for Innovation and Competition Research Paper nn. 16, 2016.

SURBLYTÉ G., *The Refusal to Disclose Trade Secrets as an Abuse of Market Dominance. Microsoft and Beyond*, Stämpfli, 2011.

SUTHARAN S., *Big Data classification: problems and challenges in network intrusion prediction with machine learning*, in *Performance Evaluation Review*, 2014.

TANG C., *The Data Industry. The Business and Economics of Information and Big Data*, Wiley & Sons, 2016.

TAPSCOTT D., *The Digital Economy: Promise and Peril In The Age of Networked Intelligence*, McGraw-Hill, 2012.

TAVANI H.T., *Philosophical theories of privacy: implications for an adequate online privacy policy*, in *Metaphilosophy*, 2017.

TAYLOR L. ET AL., *Group Privacy: New Challenges of Data Technologies*, Springer, 2017.

TENE O. - POLONETSKY J., *Big Data for all: Privacy and user control in the age of analytics*, in *Northwestern Journal of Technology and Intellectual Property*, 2013.

TORREMANS P.L.C., *Holyoak & Torremans Intellectual Property Law*, Oxford University Press, 2014.

TUCKER D.S. - WELLFORD H.B., *Big Mistakes Regarding Big Data*, in *The Antitrust Source*, 2014.

UBALDI B., *Open Government Data: Towards Empirical Analysis of Open Government Data Initiatives*, OCSE Working Paper on Public Governance, 2013.

UBERTAZZI L.C., *Introduzione al diritto europeo della proprietà intellettuale*, in *Contratto e impresa/Europa*, 2013.

VALENTE P. - IANNI G. - ROCCATAGLIATA F., *Economia digitale e commercio elettronico*, Ipsoa, Milano, 2015.

VAN ALSENOY B., *Liability under EU Data Protection Law: From Directive 95/46 to the General Data Protection Regulation*, in *Journal of Intellectual Property, Information Technology and E-Commerce Law*, 2016.

VAN DER SLOOT B. - VAN SCHENDEL S., *Ten Questions for Future Regulation of Big Data: A Comparative and Empirical Legal Study*, in *Journal Of Intellectual Property, Information Technology And E-Commerce Law*, 2016.

VAN DIJCK J., *Datafication, Dataism and Dataveillance: Big Data between Scientific Paradigm and Ideology*, in *Surveillance and Society*, 2014.

VAN HALEN ET AL., *Methodology for Product Service System Innovation*, Uitgeverij Van Gorcum, 2015.

VICTOR J.M., *The EU General Data Protection Regulation: Toward a Property Regime for Protecting Data Privacy*, in *Yale Law Journal*, 2013.

VLADECK D.C., *Charting the Course: The Federal Trade Commission's Second Hundred Years*, in *George Washington Law Review*, 2015.

WACHTER S. ET AL., *Why a right to explanation of automated decision-making does not exist in the General Data Protection Regulation*, in *International Data Privacy Law*, 2017.

WALKER G. - WEBER D., *A Transaction Cost Approach to Make-or-Buy Decisions*, in *Administrative Science Quarterly*, 2016.

WARREN S. - BRANDEIS L., *The Right to Privacy*, in *Harvard Law Review*, 2016.

WEBER R.H., *The Right to Be Forgotten: More than a Pandora's Box?*, in *Journal Of Intellectual Property, Information Technology And E-Commerce Law*, 2016.

WEBER WALLER S. - TASCH W., *Harmonizing Essential Facilities*, in *Antitrust Law Journal*, 2015.

WEISER M., *Hot Topics. Ubiquitous Computing*, in *Computer*, 2013.

WHITE G.E., *Tort Law in America: An Intellectual History*, Oxford University Press, 2016.

WHITE T., *Hadoop: The Definitive Guide*, 4th Edition, O'Reilly Media, 2015.

WIEBE A., *Protection of industrial data. A new property right for the digital economy?*, in *Journal of Intellectual Property Law & Practice*, 2017.

WILLIAMSON O.E., *Markets and Hierarchies: Analysis and Antitrust Implications*, Free Press, 2011.

WU T., *Attention Markets and the Law*, in *SSRN Library*, 2017.

WU T., *I padroni di Internet. L'illusione di un mondo senza confini*, Unwired Media, 2016.

YU P.K., *Intellectual Property and Information Wealth: Issues and Practices in the Digital Age*, Praeger, 2017.

ZECH H., *A Legal Framework for a Data Economy in the European Digital Single Market: Rights to Use Data*, in *Journal of Intellectual Property Law & Practice*, 2016.

ZECH H., *Building a European Data Economy*, in *IIC International Review of Intellectual Property and Competition Law*, 2017.

ZECH H., *Information as Property*, in *Journal Of Intellectual Property, Information Technology And E-Commerce Law*, 2015.

ZICCARDI G., *L'odio online*, Raffaello Cortina Editore, Milano, 2016.

SITOGRAFIA

https://www.agendadigitale.eu.

http://www.bollettinoadapt.it.

http://www.cloudtalk.it.

http://www.economyup.it.

http://www.ilsole24ore.com.

http://www.internet4things.it.

http://www.mo.camcom.it.

http://startupitalia.eu.

http://www.zerounoweb.it.

http://thenexttech.startupitalia.eu.

http://dimelab.us.

http://www.mantesso.com.

https://www.unirc.it.

http://www.oilproject.org.

http://it.emcelettronica.com.

https://servizi.anpal.gov.it.

http://resnovasrl.com.

http://mobile.ilsole24ore.com.

http:/www.impresaoggi.com.

http://industria40news.it.

http://www.openinnovation-platform.net.

http://www.sviluppomanageriale.it.

http://www.avioaero.com.

https://www.avioaero.com.

http://www.ilsole24ore.com.

https://www.avioaero.com.

https://www.avioaero.com.

http://www.ilsole24ore.com.

http://www.iotitaly.net.

http://www.boeingitaly.it.

http://www.thenextfactory.it.

http://www.thenextfactory.it.

http://www.stamparein3d.it.

http://www.silicon.it.

http://www.siemens.com.

http://www.industriaitaliana.it.

http://iotexpo.it.

http://w5.siemens.com.

https://www.plm.automation.siemens.com.

https://cinaoggi.it.

http://www.foxconngfo.com.

http://www.foxconndrc.com.

http://www.scmp.com.

http://www.mcar.it.

http://www.automotivemanufacturingsolutions.com.

https://ubisense.net.

http://www.impresaoggi.com.

http://www.economyup.it.

https://www.agendadigitale.eu.

https://www.worksmanagement.co.uk.