

## 1.1 | L'AMBITO DEFINITORIO

Il neologismo big data è stato spesso abusato riconducendolo a qualsiasi attività che trattasse un grande volume di dati, ingenerando alcune confusioni terminologiche.

Una delle definizioni maggiormente utilizzate negli articoli specializzati è riconducibile al rapporto McKinsey del 2011: *"Big data refers to dataset whose size is beyond the ability of typical database software tools to capture, store, manage and analyze"* (Manyika e alt. 2011, p.1). Tale definizione si svincola dal dover stabilire un "limite" numerico di volume di dati oltrepassato il quale si entra nella dimensione dei big data. Essa si pone al crocevia della convergenza tra volume di dati e capacità di immagazzinarli ed elaborarli, che devono crescere di pari passo. All'aumentare della disponibilità potenziale del volume dei dati aumenta la

capacità di immagazzinarli e la velocità di calcolo degli algoritmi che devono elaborarli. Nell'ultimo quinquennio la definizione originaria si è evoluta e arricchita al fine di comprendere in profondità questo poliedrico fenomeno. Boyd and Crawford (2012), definiscono i big data come un fenomeno culturale, tecnologico e accademico frutto dell'interazione di tre elementi: tecnologia, analisi, e mitologia, ovvero la convinzione che enormi set di dati offrano una forma più elevata d'intelligenza e conoscenza in grado di generare intuizioni, impossibili in precedenza, con l'aura di verità, obiettività e accuratezza. Gli autori aggiungono alla definizione del rapporto McKinsey l'elemento predittivo che ad oggi risulta essere una delle variabili strategiche maggiormente indagate da chi si occupa di big data.

## 1.2 | LA CRESCITA ESPONENZIALE DEI DATI

Figura I.1

### Unità di misura per i big data

FONTE: CAMICIOTTI E RACCA, 2015

<b>Byte</b>	1 byte
<b>Kilobyte</b>	1.024 byte
<b>Megabyte</b>	1.048.576 byte
<b>Gigabyte</b>	1.073.741.824 byte
<b>Terabyte</b>	1.099.511.627.776 byte
<b>Petabyte</b>	1.125.899.906.842.620 byte
<b>Exabyte</b>	1.152.921.504.606.850.000 byte
<b>Zettabyte</b>	1.180.591.620.717.410.000.000 byte

La crescita esponenziale del volume dei dati in termini di quantità e varietà è stata il risultato dell'azione congiunta di una varietà di sistemi di raccolta unita ad un drastico abbassamento dei costi e dei volumi fisici necessari ad immagazzinarli. Tale crescita è stata così smisurata da obbligare il conio di nuove unità di misura per poterla computare e quantificare. Soltanto alcuni anni addietro un gigabyte era un ordine di misura sufficiente. In breve tempo i dischi fissi dei nostri computer hanno iniziato a proporsi nell'ordine di uno e due terabyte. Nella dimensione dei big data l'ordine di misura è spesso quello dei petabyte se non degli exabyte, mentre per misurare il volume stimato

dei dati prodotti annualmente nel mondo si utilizzano i zettabyte. Nel 2013 sono stati prodotti 4,4 zettabyte (trilioni di gigabyte) di dati. Le stime prevedono per il 2020 un volume di 40 zettabyte prodotti in un anno, trecento volte quelli raccolti nel 2005. Nel 2020 ogni giorno sarà prodotto un volume di dati pari a quelli raccolti nell'intero arco dell'anno nel 2005.

I dati digitali hanno permesso una capacità di immagazzinare dati senza precedenti: il passaggio da stoccaggio analogico a digitale è stato repentino. Nel 1993 soltanto il 3 per cento dei dati immagazzinati era digitale. Nel 2000

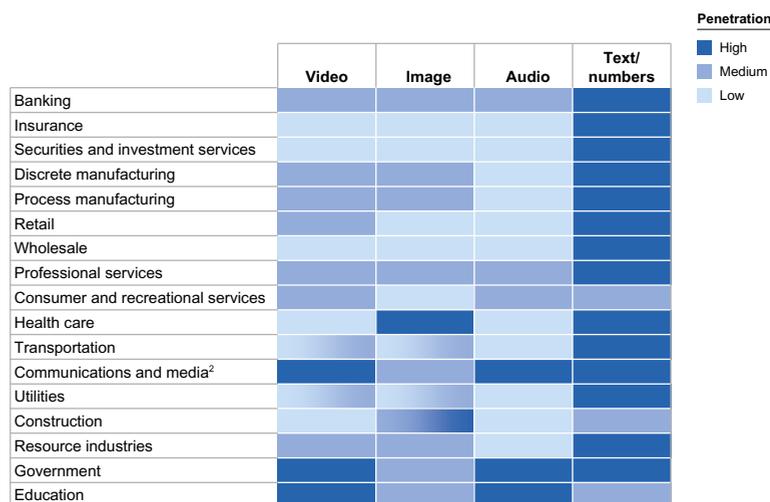
tale percentuale cresceva al 25 per cento per raggiungere il 94 per cento sette anni dopo.

Immagazzinare i dati in maniera digitale significa ricondurre immagini, video, audio, e dati numerici in formati che possano essere immagazzinati ed eventualmente in seguito elaborati. La questione è complessa: immagazzinare un dato in un formato piuttosto che in un altro permette o meno determinate elaborazioni successive. Qualora le imprese conoscano in anticipo l'utilizzo successivo dei loro dati possono rintracciare fin dall'inizio la modalità migliore di stoccaggio dei dati. Ma quasi sempre le imprese non conoscono le potenzialità future di utilizzo del loro patrimonio dei dati che se malamente stoccato potrebbe non essere utilizzabile per talune elaborazioni o confronti in futuro.

La figura mette in luce come la tipologia di dati raccolti cambia decisamente rispetto al settore di appartenenza. I dati di testo o numerici sono presenti in misura massiccia in quasi tutti i settori. I file video hanno penetrazioni molto differenti in corrispondenza dei diversi settori così come i file audio sono scarsamente presenti in molti settori.

Indipendentemente dalle loro tipologie i dati si moltiplicano ogni giorno raggiungendo volumi impressionanti. Vi sono molti esempi per rendere l'idea della smisurata mole di dati prodotti. Nel mondo ogni minuto vengono scambiati 16 milioni di messaggi, 900 persone entrano in Facebook e si spendono in media 750.000 dollari su siti di e-commerce (Barlaam 2018).

Figura I.2  
Tipologia dei dati raccolti

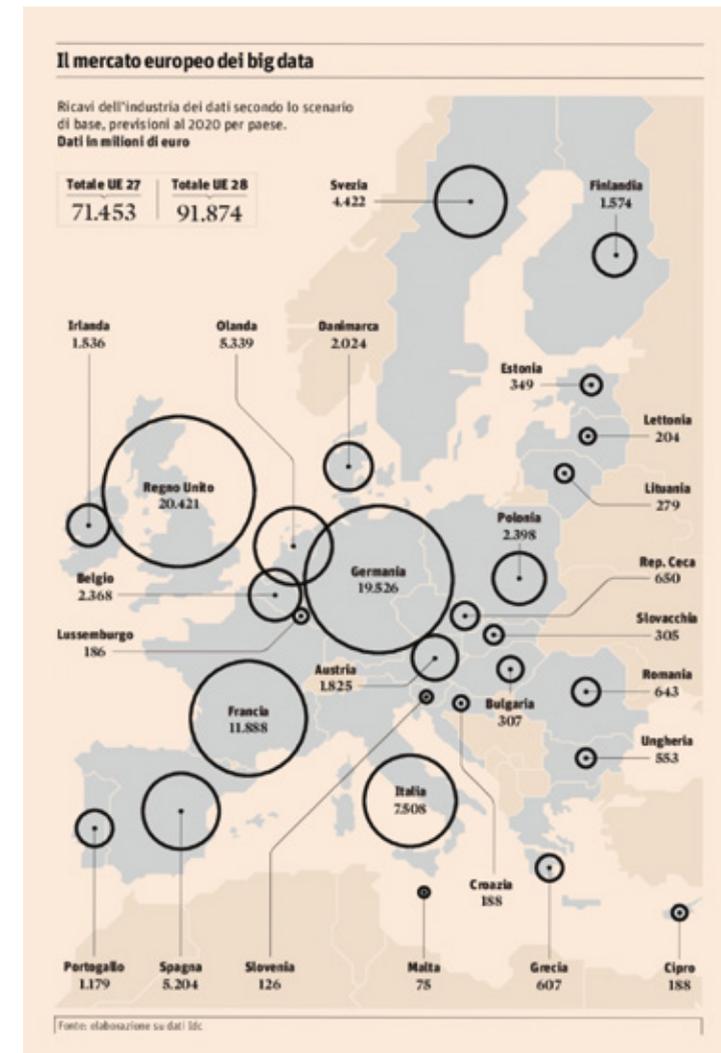


<sup>1</sup> We compiled this heat map using units of data (in files or minutes of video) rather than bytes.  
<sup>2</sup> Video and audio are high in some subsectors.  
 SOURCE: McKinsey Global Institute analysis

# 1.3 | UNA REALTÀ ECONOMICA IN ESPANSIONE

Figura I.3  
Ricavi dell'industria dei dati per paese europeo. Previsioni al 2020

FORNTE: IL SOLE 24 ORE



I dati digitali sono presenti ovunque. Non esiste un settore economico, un'impresa o una qualsiasi organizzazione che non sia in contatto con la tecnologia digitale. I big data rappresentano una potente opportunità per il mercato: i professionisti del marketing li utilizzano per indirizzare la pubblicità, i servizi assicurativi per ottimizzare le loro offerte, banchieri e operatori finanziari per interpretare il mercato. Ciò non significa che per tutte le imprese rappresenti una risorsa strategica di creazione del valore: la dimensione e il settore di appartenenza dell'impresa sono variabili rilevanti.

Per l'Italia è stato stimato che il mercato dei dati personali che comprende, produzione, raccolta, strutturazione e analisi, ha raggiunto nel 2016 il valore di 4,7 miliardi di euro. Secondo le previsioni, salirà a 7,5 miliardi nel 2020. In Europa, il valore complessivo dei ricavi dell'industria dei dati raggiungerà i 71 miliardi di euro (91 miliardi se si aggiunge il Regno Unito).

Osservando la figura emerge un ritardo della nostra economia, rispetto ai diretti concorrenti europei, a cogliere questa importante opportunità. Il totale dei ricavi stimato al 2020 per Germania e Regno Unito risulta quasi tre volte l'ammontare stimato per il nostro paese. Vi sono ampi margini di policy per avvicinare il valore di produzione della nostra industria dei dati a quello dei diretti concorrenti nei prossimi anni.

Secondo l'osservatorio Big data del politecnico di Milano, nel 2017 gli investimenti in Italia verso big data, analytics e business intelligence e data science si sono attestati a 1,1 miliardi di euro con un aumento del 21,8 per cento rispetto all'anno precedente (905 milioni di euro). L'assetto dimensionale conta in misura significativa: l'87 per cento di questi investimenti è riconducibile alle grandi imprese, in particolare alla loro corsa verso l'utilizzo di algoritmi e intelligenza artificiale. Il settore che segna la quota maggiore di investimenti è quello delle banche (28 per cento) seguito a poca distanza dalla manifattura (24 per cento) e tlc e media con una quota del 14 per cento. La crescita del mercato nazionale, sia in termini di investimenti sia di fatturato, si dimostra potente. Un segnale che i big data sono considerati un potenziale ancora in grossa parte inesplorato, nel quale le aspettative sono elevate.

Sul mercato mondiale i grandi colossi collegati ai big data hanno segnato ottime performance. Tassi di crescita del valore azionario straordinari sono stati registrati dai quattro grandi gruppi delle new economy, talmente grandi e importanti da essere identificati con un acronimo, Fang: Facebook, Apple, Netflix e Alphabet (Google).

Figura I.4  
Performance delle FANG gennaio-agosto 2017

#### FANG Stocks: YTD Performance



Confrontando il rendimento azionario delle Fang rispetto alle 500 imprese statunitensi a maggiore capitalizzazione la differenza di performance è evidente. Nell'arco di tempo considerato il rendimento dei titoli delle S&P 500 non oltrepassano la soglia del 10 per cento mentre quelli di Facebook e Netflix arrivano a lambire quella del 50 per cento, Amazon del 40 per cento e Alphabet da metà maggio supera quella del 20 per cento.

Bisogna inoltre considerare che questi quattro colossi fanno essi stessi parte dell'indice S&P 500 trainandone la crescita. Gli operatori finanziari internazionali lanciano un segnale chiaro: le loro aspettative sul futuro economico dei big data sono estremamente positive e, leggendo i dati di Wall Street per l'anno 2017, non sono state disilluse.

In generale il settore hi-tech ha registrato un aumento del 37 per cento in larga parte trainato dalle performance delle Fang.

Facebook chiude l'anno 2017 con un rialzo del titolo del 53,3 per cento, Amazon del 55,9 per cento, Google del 32,9 per cento e Apple del 46,1 per cento. Ma anche dall'altra parte del continente l'interesse per i big data si manifesta con vigore. In Cina i grandi gruppi hi-tech raddoppiano il loro valore. Le azioni di Tencent, provider di internet e proprietaria di QQ.com, il portale cinese più utilizzato sui social network e messaggistica istantanea, nel 2017 ha visto aumentare il valore delle proprie azioni del 114,5 per cento e ottime performance anche per Alibaba che ha registrato un aumento del 96,3 per cento.

Figura I.5  
Le maggiori società per capitalizzazione di Borsa. Valore in miliardi di dollari al 2017

	Market Cap	Ricavi*	Utile lordo*	Var. %
Apple	860,8	229,2	9,2	+46,1
Google	729,0	89,6	32,3	+32,9
Microsoft	660,0	96,7	3,0	+37,6
Amazon	563,5	177,1	4,9	+55,9
Facebook	514,9	40,1	5,8	+53,3
Tencent	493,8	32,9	16,7	+114,5
Alibaba	441,6	158,3	23,2	+96,3

(\*) Stime 2017 Fonte: Bloomberg

## 1.4 | ALCUNI TEMI APERTI CONNESSI AI BIG DATA

Come anticipato in precedenza il termine big data è a maglie larghe e nella letteratura accademica e stampa specializzata coinvolge questioni assai distanti tra loro. Per questo motivo abbiamo deciso di suddividere l'universo big data per otto questioni chiave ricorrenti nel dibattito internazionale. Alcune di queste sono emerse nel corso delle nostre interviste, oggetto di implementazione o valutazione di talune imprese della nostra regione. Altre abbracciano una dimensione di riflessione o grandezza di impresa per cui, al momento, non si sono ancora presentate nel panorama competitivo regionale e nazionale. Tuttavia, rappresentano le sfide future in un'arena competitiva in continua ridefinizione, nella quale le policy di accompagnamento e di indirizzo possono costituire un fattore competitivo di altissimo valore aggiunto.

### I BIG DATA SONO PORTATORI DI VERITÀ INDISCUSSE? INTELLIGENZA ARTIFICIALE VERSUS INTELLIGENZA UMANA

Nel loro articolo Boyd and Crawford (2012), si incentrano sulla dimensione "mitica" dei big data: la credenza che la connessione di intelligenza artificiale con l'enorme mole di dati sia in grado di fornire previsioni e intuizioni con obiettività e accuratezza quasi assoluta. Uno degli obiettivi, degli algoritmi allenati con i big data, è di ridurre il campo di incertezza e di rischio quando si prendono le decisioni, oltre a elaborare dei modelli predittivi. I big data non si riferiscono soltanto a set di dati molto grandi e agli strumenti e alle procedure utilizzate manipolarli e analizzarli. Essi cambiano la definizione di conoscenza innescando una svolta

computazionale nel pensiero e nella ricerca (Burkholder 1992). Non si tratta unicamente della capacità della scienza sociale computazionale di raccogliere e analizzare i dati con un'ampiezza e profondità senza precedenti (Lazer et al. 2009), quanto di un cambiamento a livello etico ed epistemologico (Boyd and Crawford 2012). In maniera provocatoria e critica Anderson pone la questione se, in quella che definisce la Petabyte Age, la nuova forza conoscitiva generata dall'analisi dei dati ponga fine alla classica ricerca teorica, rendendo obsoleti gli attuali metodi di ricerca. Secondo l'autore poiché "la correlazione sostituisce la causalità, la scienza può progredire anche senza modelli coerenti, teorie unificate o qualsiasi spiegazione meccanicistica" (Anderson 2008 p.4). Questa provocazione apre una finestra più ampia: la credenza che i big data siano portatori di verità oggettive indiscusse. Senza negare le potenzialità evidenti offerte dall'utilizzo dei big data, quando si utilizzano questi set di dati è bene usare cautela e attenzioni interpretative troppo spesso trascurate (Graham 2012). I dati personali contengono ancora molto "rumore" errori, e soprattutto vengono prodotti da persone che abitano in contesti differenti e in ambienti sociali dissimili. Ovviamente, possiamo tenere conto di tali fattori segmentando i nostri dati ma, resta comunque una parte interpretativa che rende ancora necessario il lavoro degli specialisti nel contestualizzare e offrire approfondimenti su ciò che fanno i nostri dati e soprattutto su cosa essi ci "dicono".

## IL GIANO BIFRONTE DEI BIG DATA: DAI VANTAGGI AI PERICOLI

Un volto dei big data mostra l'aspetto benigno. Da una prospettiva sociale i big data rappresentano un potente strumento per affrontare e aiutare a risolvere diversi mali della società. Offrono nuove conoscenze nella ricerca sul cancro, nella personalizzazione dei farmaci, nei cambiamenti climatici o nelle smart city. Vantaggi rilevanti sono rintracciabili nella dimensione economica: crescita delle imprese, nuove efficienze produttive, organizzative e logistiche, personalizzazione dei prodotti, aumento dell'occupazione in particolare quella giovanile e altamente scolarizzata.

L'altro volto mostra un aspetto preoccupante. La nostra identità digitale è frammentata in infinite schegge impazzite che scorrono all'interno della rete. Esse rappresentano una delle materie prime più preziose, il cui valore tenderà ad apprezzarsi notevolmente nel futuro. I big data sono percepiti come un'intrusione troppo vasta e profonda nella nostra vita personale, che riduce e condiziona il nostro grado di libertà nella sfera privata ma anche aziendale. Privacy e vantaggi economici rischiano di essere in competizione tra loro. Lawrence Lessing<sup>1</sup> (1999) si domanda se il cyberspazio debba essere regolato e come ciò possa avvenire, arrivando alla conclusione che il cyberspazio non sarà più un mondo di relativa libertà. Il rischio è che diventi un mondo di controllo perfetto nel quale le nostre azioni, desideri e identità, saranno monitorati e analizzati per esigenze di mercato. Secondo l'autore i sistemi sono regolati da quattro forze: mercato, legge, norme

sociali e architettura che nel caso dei Big data possono originare forti contrasti, in particolare tra etica e forze di mercato.

## LA QUESTIONE LEGALE DELLA PRIVACY E LE CONNESSE ASIMMETRIE COMPETITIVE

I recenti avvenimenti sul caso Facebook, coinvolto dallo scandalo di Cambridge Analytica, hanno rilanciato con forza sul piatto la questione del trattamento dei dati personali e la loro protezione. Mark Zuckerberg ha annunciato nuove regole in direzione di una stretta sulla privacy (Lops, 2018). La società, che vanta oltre due miliardi di utenti al mese, ha diffuso delle linee guida tese a rendere maggiormente informati i propri utenti sulla questione della privacy. Sarà più facile per loro individuare e gestire le proprie informazioni e i dati personali. Un aumento della protezione della privacy dell'utente potrebbe confliggere con il giro di affari della società, che tuttavia si è vista "costretta" a cambiare direzione dopo le ripercussioni negative del titolo in borsa a seguito dello scandalo Cambridge Analytica. Ma non è stata la borsa l'unico strumento di persuasione.

Parallelamente si è creato il fronte #delete Facebook che ha allargato in poco tempo i propri sostenitori, capeggiati da importanti nomi dell'industria, della cultura e dello spettacolo.

Se negli Stati Uniti tali iniziative sono riconducibili soprattutto a singole imprese, l'Unione europea si è recentemente mossa con forza per affrontare in un quadro regolatorio la tematica della privacy legata alla raccolta dei dati. Mediante il regolamento generale sulla protezione dei dati (*GDPR, General Data*

*Protection Regulation*)<sup>2</sup> la Commissione europea mira a irrobustire e rendere più omogenea la protezione dei dati personali dei propri cittadini. Il testo, adottato il 27 aprile 2016, ed entrato in vigore il 25 maggio del 2016, inizierà ad avere efficacia il 25 maggio 2018.

Il GDPR è una delle parti di regolamento più complesse che l'Unione Europea abbia mai prodotto (The Economist, 2018 a). Il Garante della privacy, in Italia, si è mosso fin dallo scorso anno<sup>3</sup>, organizzando una serie di incontri con imprese e pubblica amministrazione finalizzati a supportare l'implementazione del Regolamento individuando soluzioni per gestire al meglio la transazione verso le nuove regole. Esse sono concepite per migliorare il modo in cui i dati sono raccolti, archiviati e utilizzati, consegnando alle persone un maggiore controllo rispetto alle loro informazioni e obbligando le imprese a gestire i dati con maggiore attenzione. Ad esempio, la richiesta per il consenso per la raccolta ed elaborazione dei dati personali non potrà essere ambigua e non vi potranno essere clausole catch-all nascoste. Le persone potranno richiedere una copia dei dati in loro possesso (data portability), chiedere di correggere le informazioni (right to rectification) e richiedere la cancellazione dei propri dati (right to be forgotten). Le aziende che ignorano costantemente le norme del regolamento possono pagare multe fino a 20 milioni di euro o il 4 per cento delle vendite annuali globali (The Economist, 2018). Inoltre, le imprese devono assicurarsi che le imprese da cui ricevono i dati personali e quelle a cui li inviano siano in conformità con il regolamento. L'aumento della complessità per la gestione dei

dati personali potrebbe condurre molte imprese a esternalizzare questa funzione. Secondo alcune stime circa il 60 per cento delle imprese non sono ancora pronte per adeguarsi al regolamento, in particolare quelle di piccole dimensioni che non dispongono delle risorse necessarie per organizzarsi nei termini previsti del regolamento.

Infine, le complesse maglie normative del nuovo regolamento, di non facile e immediata applicazione, renderanno più onerosa e meno vasta la raccolta e stoccaggio di dati personali da parte delle imprese. Ciò potrebbe rappresentare un ostacolo, in Europa, per lo sviluppo di servizi di intelligenza artificiale, che necessitano di una grande quantità e varietà di dati per essere perfezionati, oltre a limitare tutte le altre applicazioni tese ad accrescere l'efficienza e il business delle imprese. Si verrebbe così a creare un'asimmetria competitiva su scala globale che penalizzerebbe le imprese europee rispetto a quelle americane e asiatiche. È perfino ipotizzabile, nel futuro, che taluni governi potranno emanare regolamenti sulla protezione dei dati personali a favore delle proprie imprese, rendendo di fatto non disponibili i dati oltre i propri confini.

## LE COMPETENZE SPECIALIZZATE: UNA RISORSA SCARSA E APPETIBILE

La comparsa dei big data nel panorama economico ha generato nelle imprese una corsa per assicurarsi i ricercatori e i laureati con le adeguate competenze. Improvvisamente tali competenze sono diventate una risorsa estremamente rara quanto preziosa. Nel big data

<sup>1</sup> Fondatore dello Stanford Center for Internet and Society.

<sup>2</sup> Regolamento (UE) 2016/679 del 27 aprile 2016 relativo alla protezione delle persone fisiche con riguardo al trattamento dei dati personali, nonché alla libera circolazione di tali dati e che abroga la direttiva 95/46/CE (regolamento generale sulla protezione dei dati).

<sup>3</sup> Newsletter dell'Autorità Garante, n.428 del 25 maggio 2017.

l'offerta di lavoro è cresciuta in maniera così repentina da ampliare di anno in anno la distanza tra offerta e domanda di lavoro, sospinta dalla somma di due effetti che possiamo definire *effetto espansione* ed *effetto allargamento*. Al primo effetto è riconducibile la crescita dell'offerta di lavoro da parte di imprese che già avevano al proprio interno queste competenze (ingegneri informatici, elettronici fisici etc..) e, spinte dalle potenzialità offerte dai big data, hanno deciso di potenziare la propria squadra di analisi e ricerca. L'*effetto allargamento* descrive un fenomeno nuovo. Imprese di settori che storicamente non hanno mai utilizzato tali competenze, quali il settore assicurativo, della distribuzione o finanziario, che necessitano ora di assumere competenze specializzate per gestire, soprattutto nella fase di elaborazione, il potenziale offerto dall'enorme mole di dati di loro proprietà. Risorse umane che prima erano contese unicamente da imprese di settori contigui si trovano a essere appetibili per imprese di qualsiasi settore economico, produttivo e di servizi.

Nel panorama internazionale è il campo dell'AI quello che registra la maggiore corsa frenetica per assicurarsi i migliori talenti. Nell'AI le risorse umane rappresentano un bene più scarso rispetto ai dati o alla potenza di calcolo. La domanda di questi "fabbricanti" di intelligenza artificiale, in grado di applicare tecniche di apprendimento automatico a gigantesche serie di dati, è cresciuta in maniera esponenziale oltrepassando agevolmente il numero di studenti in possesso di tali competenze.

I sistemi IA possono essere considerati come "idioti dotti" (The Economist 2017 a) e, per evitare disastri, devono essere utilizzati da personale estremamente competente. Assumere personale in grado di gestirli adeguatamente può rivelarsi fondamentale per la sopravvivenza di un'impresa che abbia investito pesantemente in AI. Per questo motivo si è innescata una vera caccia alle risorse umane. Le imprese saccheggiano i dipartimenti accademici per assumere professori e studenti ancor prima che finiscano il loro percorso di studio<sup>4</sup>. Persino le conferenze accademiche si trasformano in luoghi nei quali acquisire i talenti. Negli Stati Uniti i migliori reclutatori sono celebri ex professori che mantengono stretti contatti con l'università<sup>5</sup> e possono attrarre ricercatori al loro fianco offrendo loro lautissimi stipendi combinati all'accesso ai dati proprietari dell'impresa necessari per le loro pubblicazioni. Talvolta questa corsa frenetica raggiunge toni elevati laddove imprese comprano altre imprese unicamente per assicurarsi il loro patrimonio di risorse umane. Correva l'anno 2014, quando Google spese circa 500 milioni di dollari per acquisire la startup inglese DeepMind, che non registrava ricavi né aveva in portafoglio prodotti commerciabili. Lo scopo principale di Google era acquisire una squadra di ricercatori ad alta capacità di conoscenza.

#### APERTURA VERSUS CHIUSURA

*Indoor knowledge*, oppure *outdoor knowledge*? Il dilemma della conoscenza ristretta o condivisa si ripropone con forza nei big data. Se i big data rappresenteranno la scommessa strategica dei prossimi anni è comprensibile

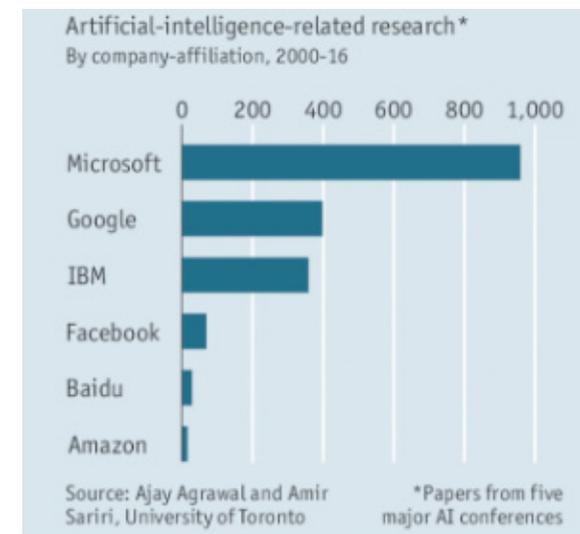
che le imprese abbiano un indubbio vantaggio a custodire gelosamente i loro *secrets of knowledge* all'interno delle loro mura. In tal caso la condivisione di conoscenza e delle banche dati è marginale se non assente. Tuttavia, vi possono essere alcune forze che spingono nella direzione contraria: il reclutamento di ricercatori, l'ampiezza e la trasversalità della conoscenza da governare e la dimensione dell'impresa.

In merito al primo e secondo punto i colossi americani perseguono filosofie diverse: alcune più votate all'*outdoor knowledge*, altre che sposano rigidamente l'*indoor knowledge*.

Imprese come Microsoft e IBM, investono molto nella ricerca nel campo dell'AI e i loro ricercatori pubblicano numerosi articoli. Solitamente non richiedono ai loro ricercatori di applicare per forza le loro scoperte ad attività di business (The Economist 2018a). All'estremo opposto vi sono gruppi come Apple e Amazon, nei quali la ricerca è fortemente orientata al mercato e si attendono soluzioni di business dal lavoro dei propri ricercatori. Idealmente nel mezzo possiamo posizionare Google e Facebook i quali ricercatori, pur dovendo impegnarsi su percorsi di ricerca orientati al business hanno alcuni gradi di libertà e condivisione della ricerca.

Figura I. 6  
Numero di articoli pubblicati dei ricercatori in impresa

Fonte: THE ECONOMIST



<sup>4</sup> Esempio il caso della Carnegie Mellon University (CMU), un'istituzione pionieristica nell'IA il cui dipartimento di robotica è stato saccheggiato da Uber nel 2015.

<sup>5</sup> Ad esempio Yann LeCun di Facebook. Nel 2012 ha fondato, presso la New York University il Centro della Scienza dei Dati. Nel dicembre 2013 è diventato il primo direttore del Facebook AI Research interrompendo il lavoro presso la New York University nei primi mesi del 2014. Oppure anche il caso di Geoffrey Hinton di Google.

Si crea un trade-off tra bisogno di segretezza dell'impresa e necessità di reclutamento di personale scarsamente presente sul mercato. Le imprese che permettono ai loro ricercatori di condividere all'esterno i percorsi di ricerca, esercitando una minore pressione sulle attese economiche di risultati, riusciranno più facilmente ad attrarre talenti. La condivisione esterna dei propri risultati di ricerca si traduce in una maggiore possibilità di carriera del ricercatore e quindi può essere una condizione che le risorse più competenti valutino attentamente prima di firmare un contratto di lavoro. Il conflitto tra apertura e chiusura dei risultati della ricerca coinvolge anche colossi cinesi come Baidu che nel 2013 e quest'anno ha aperto, nella Silicon Valley, due laboratori di ricerca incentrati sull'AI. I ricercatori occidentali li valutano molto positivamente tuttavia preferiscono

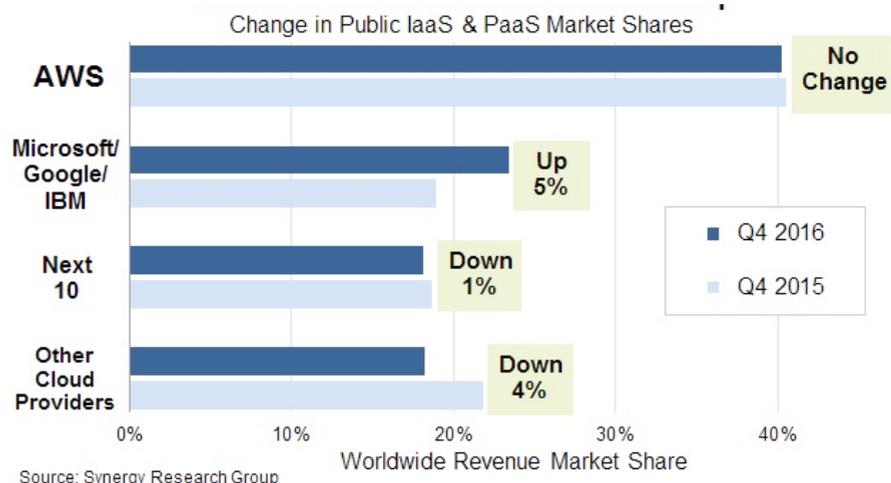
ancora lavorare per i giganti americani, anche in virtù della loro relativa trasparenza (the Economist 2018 a).

**DAL CENTRO ALLA PERIFERIA DEI DATI: IL RUOLO DELLA IOT**

I big data spingono verso le concentrazioni, delle imprese, della conoscenza e dei dati. L'avvento del cloud negli anni 2000, ha spinto verso una nuova centralizzazione con nuove imprese leader come Amazon Web Services (AWS), Microsoft Azure e Google Cloud. Secondo il rapporto di Synergy Research del febbraio 2017, Amazon Web service detiene il 40 per cento del mercato dei servizi cloud, mentre Microsoft, Google e IBM, insieme, controllano il 23 per cento (Synergy Research Group 2017).

Figura I.7  
**Quote nel mercato dei servizi pubblici cloud. IaaS e PaaS, 2016**

Fonte: THE ECONOMIST

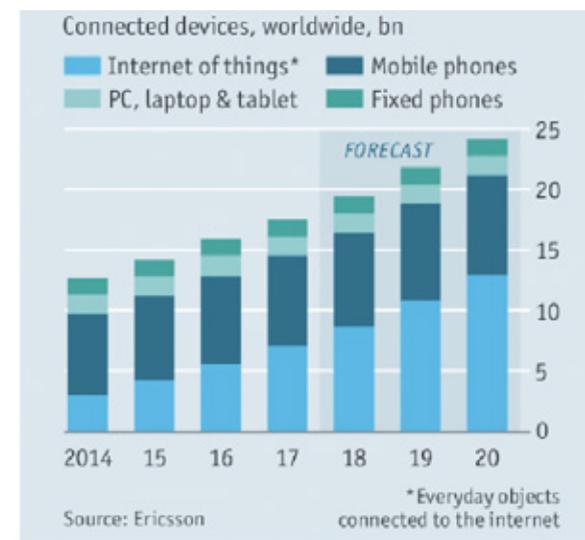


Dalla figura emerge come in un anno sia aumentato l'indice di concentrazione nel settore. Se da un lato Amazon Web Services (AWS) mantiene stabile la sua quota dominante del mercato dei servizi cloud, crescono del 5 per cento le quote dei diretti rivali, Microsoft, Google e IBM, che guadagnando terreno a spese degli operatori più piccoli sul mercato facendo crescere la quota dei primi quattro operatori al 63 per cento. In poco tempo il cloud è diventato il "luogo" dove convergono migliaia di miliardi di dati che vengono giornalmente immagazzinati. Tuttavia, si sono alcuni segnali di uno spostamento dei dati dalla centralizzazione del cloud verso il "bordo" di reti e dispositivi intelligenti in virtù dell'accresciuta presenza dell'internet of things (The Economist 2018 b).

La raccolta dei dati aumenta in maniera esponenziale grazie ai dispositivi connessi che a loro volta si moltiplicano in serie geometrica assimilando qualsiasi tipologia di informazioni che poi sono inviati a coloro che devono analizzarli. La crescita tecnologica rende i dispositivi quali gli smartphone o le macchine industriali sempre più intelligenti dotate di potenti processori, in grado di affrontare i problemi informatici che solo alcuni anni fa dovevano essere delegato ad un server specifico. Queste nuove capacità tecnologiche rendono possibile una più ampia distribuzione dei dati informatici. La quota degli oggetti collegati ad internet (internet of things) aumenta più che proporzionalmente nel corso degli anni, raggiungendo, nelle stime al 2020, oltre la metà dei dispositivi connessi.

Figura I.8  
**Dispositivi connessi nel mondo, dati in miliardi**

Fonte: THE ECONOMIST



Parte della competizione sull'acquisizione di dati si sposterà dal centro "cloud" verso la periferia dove risulteranno vincenti quelle imprese che per prime la riusciranno a colonizzarla.

La dimensione dell' "internet of things" (IoT), giocherà un ruolo fondamentale in questa battaglia. Bisogna comprendere se questa battaglia avvantaggerà gli attuali *incumbent* oppure costituirà una porta aperta per eventuali *entrant*. Pertanto, ci si domanda se gli attuali grandi fornitori di cloud come Amazon Web Services (AWS) o Microsoft riusciranno ad estendere le loro maglie per governare anche la periferia. Ad esempio, alcuni ipotizzano che uno dei motivi per cui Amazon, nel 2017, ha acquisito, per circa 14 miliardi di dollari la catena di negozi di generi alimentari Whole Foods, fosse per appropriarsi dei loro data center locali (The Economist 2018 b).

Oltre alla variabile tecnologica, lo spostamento dei dati verso il confine della periferia è anche sospinto dall'accresciuto numero delle imprese che, in virtù di comportamenti strategici o di una pressione delle normative sulla gestione dei dati personali, preferiscono mantenere al proprio interno i dati.

Sicuramente il dominio centrale della "nuvola" resterà saldo. Tuttavia i nuovi scenari offerti

dall'IoT e dalle nuove regolamentazioni sui dati personali, potrebbero rafforzare l'importanza della raccolta e stoccaggio di dati nel "bordo". Si aprirebbe così la strada per nuove ed innovative imprese in un mercato attualmente controllato e dominato da pochissime imprese.

### LA BATTAGLIA SULL'INTELLIGENZA ARTIFICIALE

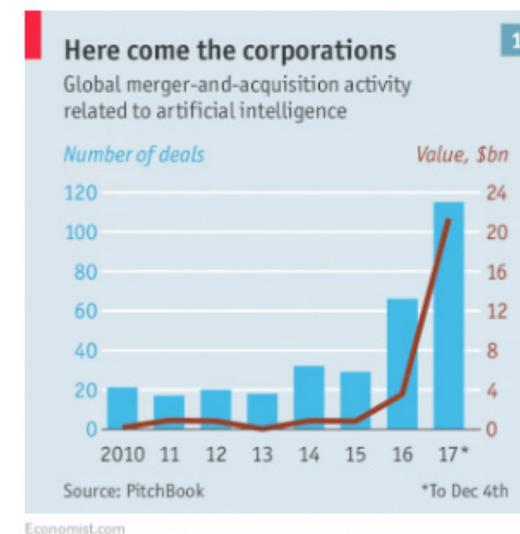
La tematica della nuova competizione sul terreno dell'intelligenza artificiale richiama inevitabilmente la questione dimensionale. I grandi gruppi globali si stanno muovendo con decisione verso questa nuova frontiera. Gli investimenti in venture-capital in AI, nei primi nove mesi del 2017 hanno totalizzato un valore di 7,6 miliardi di dollari (The Economist 2017 b).

Nel corso del 2017 si sono succedute acquisizioni e fusioni con un vigore senza precedenti. Pur considerando la difficoltà di separare gli investimenti delle imprese tecnologiche nell'IA da altri tipologie, secondo PitchBook, le imprese, nel corso del 2017, hanno realizzato fusioni e acquisizioni legate all'IA per un valore di 21,3 miliardi di dollari, cinque volte maggiore dell'anno precedente e 26 volte il valore di due anni prima.

Figura I.9

### Operazioni di fusioni e acquisizioni in attività AI, per numero e valore

FONTE: THE ECONOMIST



In particolare il machine learning è il settore dell'AI che più coinvolge le operazioni di M&A (Merge&Acquisition). Il dato che più impressiona è la crescita enorme del valore medio delle operazioni di M&A. Rispetto al 2016 il numero delle M&A è poco meno del doppio mentre il valore totale delle operazioni coinvolte è oltre cinque volte. Le acquisizioni e fusioni diventano sempre più corpose e ci si domanda se queste dinamiche aumenteranno il potere di mercato del *incumbent* esistenti. I gladiatori di questa battaglia sono nomi noti, Alphabet è considerato uno dei maggiori protagonisti. Per molti anni ha realizzato notevoli profitti dall'intelligenza artificiale e la sua squadra di ricerca annovera molti dei ricercatori più stimanti in questo campo (The Economist 2017 a). Vi sono poi Amazon, Apple, Facebook e

Microsoft e si stanno interessando alla IA colossi cinesi come Alibaba e Baidu.

### LA QUESTIONE DIMENSIONALE ED IL "RITORNO" DEGLI INCUMBENT

La battaglia sull'intelligenza artificiale coinvolge una tematica più ampia: la questione dimensionale. Ci si domanda se la nuova competizione nella AI, e più in generale nei big data, sia così selettiva in termini dimensionali, che nel futuro al mondo solo pochi colossi avranno le dimensioni adeguate per gestire tali complessità competitive: un mondo di *incumbent* senza *entrant*.

I colossi che operano nei big data e nella AI, vantano la possibilità di poter disporre di enormi volumi di dati, incredibili capacità di

calcolo e abbondanza di ricercatori specializzati, solitamente i più quotati sul mercato. Il mix di questi tre elementi consegna un indiscusso vantaggio competitivo rispetto alle imprese che a fatica governano uno o al massimo due di questi tre ingredienti del successo.

Tuttavia, vi sono alcune questioni che possono minare la capacità degli attuali *incumbent* di restare tali nel tempo. Anzitutto è necessario comprendere quale sarà la quota mondiale di dati che i giganti della tecnologia riusciranno a controllare. Come anticipato nella tematica relativa alla regolamentazione della privacy, vi sono pressioni dagli utenti privati e regolazioni dalle autorità pubbliche che renderanno più complesso e costoso per le imprese ottenere i dati: ciò comporterà che il valore dei dati personali crescerà nel tempo. In secondo luogo, non è scontato che le performance nell'AI dipenderanno sempre più dalla quantità di dati disponibili (The Economist 2017 b). Oggi le

enormi quantità di dati servono per addestrare le macchine intelligenti: chi possiede più dati in termini di quantità, qualità e varietà, meglio le addestra. Una visione competitiva dell'IA potrebbe presupporre un ambiente di simulazioni nel quale le macchine si auto-addestrano utilizzando dati sintetici o ambienti virtuali. In tal caso possedere una grande disponibilità di dati vedrebbe limitato il suo potere di mercato in merito all'innovazione e sviluppo delle AI. L'ultima questione, trattata in precedenza, è relativa all'open knowledge, al fine di reclutare le migliori competenze. Alcuni giganti come Google e Facebook hanno reso disponibili alcune librerie software a sviluppatori esterni e questo potrebbe comportare un abbassamento delle barriere all'entrata. Tuttavia, non appena ci si addentra nel campo di algoritmi strategici la loro propensione alla condivisione si riduce notevolmente a favore dei *secrets of knowledge*.