



Thomas H. Davenport

BIG DATA @L LAVORO

Sfatare i miti, scoprire le opportunità



FRANCOANGELI

NEO
MANAGEMENT

Informazioni per il lettore

Questo file PDF è una versione gratuita di sole 20 pagine ed è leggibile con



La versione completa dell'e-book (a pagamento) è leggibile con Adobe Digital Editions. Per tutte le informazioni sulle condizioni dei nostri e-book (con quali dispositivi leggerli e quali funzioni sono consentite) consulta [cliccando qui](#) le nostre F.A.Q.





SCENARI E STRUMENTI DELL'INNOVAZIONE DIGITALE

“Neo” è il primo progetto editoriale crossmediale di saggistica che narra e spiega l’innovazione digitale attraverso l’innovazione digitale stessa. Per ogni titolo un testo cartaceo con un ebook e una piattaforma digitale di conversazione (<http://blog.francoangeli.it/neo/>), alimentati e aggiornati progressivamente dall’autore e arricchiti dagli utenti. Un processo narrativo e divulgativo continuo.

La collana affronta l’innovazione digitale nel suo complesso, come fenomeno culturale, scientifico e sociale e, contemporaneamente, ne approfondisce, in maniera verticale e specialistica, il particolare.

“Neo” è un prefisso da apporre alle varie aree trattate: impresa, management, comunicazione, scienze sociali, sanità, pubblica amministrazione. Per ogni area tematica, la collana raccoglie le opere di esperti, nazionali e internazionali. Lo scopo è di fornire la conoscenza degli strumenti per cogliere le opportunità che l’innovazione digitale apre e di tracciare il percorso cognitivo, in forma rigorosa e divulgativa, per comprendere i nuovi universi concettuali e operativi.

Un punto di incontro per tutti coloro che vogliono capire e confrontarsi in questa “neo-società”.

Direttore

Daniele Chieffi – Head of Media Relations Online ENI

Board scientifico

Paola Bacchiddu – Giornalista

Davide Bennato – Sociologo, Università di Catania

Giovanni Boccia Artieri – Sociologo, Università di Urbino

Luca De Biase – Direttore *Nova Il Sole 24 Ore*

Ruggero Eugeni – Direttore Almed, Scuola di Alta Formazione dell’Università Cattolica di Milano

Cristiano Habetswallner – VP Sponsorship Management Gruppo Telecom Italia

Sergio Maistrello – Giornalista e scrittore

Roberta Milano – Direttrice scientifica Italia di BTO (Buy Tourism Online)

Eugenio Santoro – Ricercatore, Istituto Mario Negri Milano

I lettori che desiderano informarsi sui libri e le riviste da noi pubblicati possono consultare il nostro sito Internet: www.francoangeli.it e iscriversi nella home page al servizio “Informatemi” per ricevere via e.mail le segnalazioni delle novità o scrivere, inviando il loro indirizzo, a “FrancoAngeli, viale Monza 106, 20127 Milano”.



Thomas H. Davenport

BIG DATA @L LAVORO

Sfatare i miti, scoprire le opportunità

FRANCOANGELI

NEO
MANAGEMENT

Progetto grafico di copertina: Elena Pellegrini

Immagine di copertina: *Bits Cubes. Digital Data Isolated* © Dreaming Andy – Fotolia.com

Titolo originale: *Big Data @ Work. Dispelling the Myths, Uncovering the Opportunities*,
Harvard Business School Publishing, 60 Harvard Way, Boston, Massachusetts.

Copyright © 2014 Harvard Business School Publishing Corporation

Published by arrangement with Harvard Business Review Press

Traduzione dall'inglese di Mauro Formaggio

1ª edizione. Copyright © 2015 by FrancoAngeli s.r.l., Milano, Italy.

L'opera, comprese tutte le sue parti, è tutelata dalla legge sul diritto d'autore. L'Utente nel momento in cui effettua il download dell'opera accetta tutte le condizioni della licenza d'uso dell'opera previste e comunicate sul sito www.francoangeli.it.

Indice

Ringraziamenti	pag.	9
1. Perché i big data sono importanti per voi e la vostra organizzazione	»	11
1.1. Oltre la moda dei big data	»	12
1.2. Chi siete?	»	14
1.3. Decostruire l'espressione "big data"	»	15
1.4. I big data dureranno?	»	18
1.5. Che cosa c'è di nuovo, dal punto di vista del management?	»	23
1.6. Nuovi orientamenti organizzativi stimolati dai big data	»	26
1.7. Nuove opportunità dai big data	»	30
1.8. Ciò che non sappiamo e continueremo per qualche tempo a non sapere	»	33
1.9. Il contenuto del libro	»	35
2. Come i big data cambieranno il vostro lavoro, la vostra azienda e il vostro settore	»	38
2.1. Quattro scenari futuri	»	39
2.1.1. Big data e viaggi d'affari	»	39
2.1.2. Uno scenario di big data per l'energy management	»	41
2.1.3. Uno scenario di big data per la video analytics nel retail	»	43

2.1.4. Uno scenario di big data per la home educa- tion	pag. 45
2.2. Tradurre gli scenari in realtà	» 46
2.3. Settori adatti ai big data	» 47
2.4. Funzioni organizzative adatte ai big data	» 54
2.5. In sintesi	» 59
3. Sviluppare una strategia	» 61
3.1. Qual è il vostro obiettivo?	» 61
3.1.1. Riduzione dei costi	» 62
3.1.2. Risparmi di tempo	» 64
3.1.3. Nuove offerte	» 66
3.1.4. Ausilio alle decisioni interne	» 68
3.2. Discovery e produzione	» 71
3.2.1. Discovery	» 71
3.2.2. Produzione	» 72
3.3. Progettare nuove iniziative	» 74
3.4. Chi partecipa a che cosa?	» 76
3.5. Da dove iniziare?	» 77
3.6. Con quale rapidità muoversi?	» 79
3.6.1. Stile cauto	» 80
3.6.2. Stile moderatamente aggressivo	» 81
3.6.3. Stile aggressivo	» 82
4. Il lato umano dei big data	» 84
4.1. Che cosa c'è di nuovo nella figura del data scientist?	» 85
4.2. Il modello classico di data scientist	» 85
4.2.1. Hacker	» 87
4.2.2. Scienziato	» 88
4.2.3. Consulente di fiducia	» 90
4.2.4. Analista quantitativo	» 91
4.2.5. Esperto di business	» 94
4.3. Data scientist orizzontali e verticali	» 94
4.4. L'apporto del team	» 96
4.5. Dove trovare data scientist?	» 98

4.5.1. Università	pag. 98
4.5.2. Fonti non accademiche	» 99
4.6. Retention dei data scientist	» 100
4.7. Competenze per i manager	» 102
4.8. Capacità umane e big data	» 105
5. Tecnologie per i big data	
Scritto insieme a <i>Jill Dyché</i>	» 108
5.1. Che cosa c'è di veramente nuovo nella tecnologia per i big data?	» 109
5.2. Lo stack dei big data	» 112
5.3. Integrazione delle tecnologie per i big data	» 118
5.4. Che cosa fanno oggi molte grandi aziende	» 120
5.5. Coesistenza di diversi elementi	» 122
6. Che cosa occorre per avere successo con i big data	» 126
6.1. DELTA rivisitato	» 126
6.2. Data: i dati	» 127
6.3. Enterprise: l'impresa	» 129
6.4. Leadership	» 130
6.5. Target: gli obiettivi	» 133
6.6. Analysts: gli analisti	» 134
6.7. Altri fattori da considerare per avere successo con i big data	» 135
6.7.1. Cultura	» 136
6.7.2. Inclusione dei big data nei principali processi di business	» 138
7. Che cosa potete imparare dalle start-up e dalle aziende dell'online	» 141
7.1. Esempi da seguire	» 142
7.1.1. Utilizzare i big data per l'innovazione di prodotto e di servizio	» 142
7.1.2. Lavorare agli strumenti, non solamente alle applicazioni	» 143

7.1.3. Dare ai data scientist i capi adatti	pag. 144
7.1.4. Badare alla produttività del lavoro sui big data	» 146
7.1.5. Contribuire al patrimonio comune	» 147
7.1.6. Ricordarsi che non si è mai troppo veloci!	» 147
7.1.7. Approfitte di ciò che è gratuito o costa poco	» 149
7.1.8. Sperimentare su grande scala	» 150
7.1.9. Promuovere una più stretta collaborazione	» 151
7.2. Errori da evitare	» 152
7.2.1. Non condividere i dati con i clienti	» 152
7.2.2. Raccogliere dati senza uno scopo preciso	» 153
7.2.3. Parlare troppo di tecnologia	» 155
7.2.4. Indulgere nelle false promesse	» 156
7.2.4. Concludendo	» 157
8. Che cosa potete imparare dalle grandi aziende. Big data e Analytics 3.0	» 158
8.1. Quale novità?	» 159
8.2. Esempi di obiettivi per i big data in grandi aziende	» 161
8.3. Integrazione delle strutture organizzative e delle competenze	» 163
8.4. La value proposition dei big data	» 167
8.5. Redditività dell'investimento	» 168
8.6. Automazione di processi esistenti	» 171
8.7. Le nuove opportunità	» 172
8.8. L'ascesa dell'Analytics 3.0	» 173
8.9. In sintesi	» 180
Appendice. Valutazione dell'adeguatezza ai nuovi progetti centrati sui big data	» 183
L'autore	» 187

Ringraziamenti

Vorrei essere così brillante da presentare in questo libro idee originali, solamente mie. Ma così non è. Mi trovo quindi a dipendere dalla cortesia di persone competenti che mi tengono al corrente di ciò che stanno facendo con i big data. Sono perciò grato a tutti coloro che hanno condiviso con me le loro esperienze e le loro riflessioni. Dietro a ognuna delle numerose vicende presentate nel libro c'è una persona generosa.

Sono anche grato ai miei amici del SAS Institute, che ha sponsorizzato gran parte della mia attività di ricerca in materia di analytics. Nel libro sono descritti due degli studi che rientrano in tale attività. Mike Bright è il mio riferimento principale in SAS, dove il mio contatto iniziale è stato con Scott VanValkenburg. Dopo una collaborazione quasi decennale, entrambi continuano a essere miei buoni amici e consiglieri. Meritano un ringraziamento, in SAS, anche Ken Blank, Jim Davis, Carl Farrell, Jim Goodnight, Deb Orton, Adele Sweetwood e tanti altri che sono troppo numerosi per essere citati individualmente.

L'attività di ricerca sulla quale si basa questo libro si è svolta in parte grazie all'aiuto dell'International Institute for Analytics (IIA), fondato da me e da Jack Phillips. Grazie quindi a Jack, a Katherine Busey, a Sarah Gates, a Callie Youssi e a tutti gli altri colleghi dello IIA per l'aiuto che mi hanno fornito.

Sono da parecchi anni senior adviser presso Deloitte Analytics, dove Jane Griffin mi ha sempre dimostrato sostegno e amicizia. Ringrazio anche Forrest Danson, Wendy DeHoef, Kelly Nelson, Tim Phillipps e Fred Roffman – tutti colleghi a Deloitte Analytics – per la loro costante assistenza. In Deloitte, Marcus Shingles ha diretto un progetto di ricerca sui big data presso la Grocery Manufacturers Association al quale ho partecipato anch'io. Nel libro sono presenti alcune delle idee emerse in quell'occasione.

Infine, durante la stesura del libro ho iniziato un progetto di ricerca in materia di esplorazione (*data discovery*) e alcune indicazioni che ne andavo

traendo sono confluite nel testo, in corso d'opera. Teradata Aster ha finanziato il progetto. Ringrazio perciò Tasso Argyros e Mary Gros che mi hanno dato accesso ai dati sui loro clienti e sui loro prodotti.

Collaboro inoltre con diverse aziende e organizzazioni in qualità di consulente o di ricercatore e ho imparato molto da ognuna di esse. Vorrei qui citare, in ordine alfabetico, First Analytics, MarketShare, Medidata Solutions, il MIT Center for Digital Business, Mu Sigma, Real Impact, Signals Intelligence Group, Via Science. A tutte queste organizzazioni dico sempre che il mio intento principale è quello di imparare da esse. Tutte hanno risposto positivamente, nei fatti.

Altre persone che hanno avuto una parte rilevante nella realizzazione di questo libro lavorano alla Harvard Business Review Press. Lì, Melinda Merino ha già curato l'edizione di altri miei libri e anche in questo caso è stato un piacere lavorare insieme a lei. Non ricordo con sicurezza, tuttavia penso che l'idea di scrivere il libro sia stata sua. Certamente Melinda mi ha dato consigli preziosi in corso d'opera, per rendere il libro più interessante e più incisivo.

Nei miei volumi precedenti ho mancato di ringraziare Julie Devoll, che da tempo è la mia publicist alla Harvard Business Review Press. Quando scrivo i ringraziamenti, solitamente non so ancora chi sarà a occuparsi della promozione di un mio libro ma Julie mi ha assicurato che in questa occasione sarà lei a fornirmi i suoi impeccabili servizi. Chiunque sentirà parlare di questo libro, o lo leggerà, lo farà quindi probabilmente anche per merito di Julie.

Tracy Williams ha coordinato il team di marketing, mentre Allison Peter si è occupata con bravura e competenza dell'editing e della grafica. Ringrazio anche Julia Kirby, amica di lunga data, che è editor presso la *Harvard Business Review*, per avermi aiutato a chiarire le mie idee sull'argomento in un paio di articoli che ho scritto per la rivista.

Mia moglie, Jodi Davenport, mi ha incoraggiato a scrivere un libro sui big data assai prima che io vi ponessi mano. I miei figli, Hayes e Chase, hanno anch'essi un certo interesse per l'argomento – Hayes per quanto riguarda il settore dell'entertainment, Chase in relazione all'ambito educativo. Grazie a entrambi per i loro consigli.

Ho scritto più libri di quanti siano i miei parenti e ora mi trovo a ripetere dediche già fatte in precedenza. Ho già dedicato a mia suocera, Helen Kubik, il libro *Competing on Analytics*. Non solo lei ha gradito più di qualunque altro questo genere di attenzione, ma quel libro ha avuto un notevole successo. La dedica perciò è a Helen, ancora una volta, con l'auspicio che lei ripeta anche in questa occasione la sua magia di allora!

Perché i big data sono importanti per voi e la vostra organizzazione

I “big data” sono senza dubbio “big”, ma sono anche altro. L’espressione “big data” non è del tutto calzante, perché in effetti si riferisce, più in generale, a tutti quei dati che non rientrano nei contenitori abituali: dati che hanno una mole o, come anche si dice, un “volume” troppo grande per poter essere gestiti in un unico server; dati che sono troppo poco strutturati per essere trattati in un database ordinato per righe e colonne; dati che si presentano in un flusso continuo che mal si adatta a un data warehouse statico. Insomma, l’espressione “big data” concentra l’attenzione sulla grande mole, ma è l’assenza di struttura l’aspetto più significativo che caratterizza questi dati.

I libri su questo argomento iniziano generalmente citando la quantità di dati presenti nel mondo. Sono numeri e raffronti ben noti: l’azienda media ha oggi una mole di dati 427 volte maggiore di quella registrata nella Biblioteca del Congresso degli Stati Uniti; Facebook ha più dati fotografici di quanti siano i pixel mai elaborati dalla Kodak; vengono registrati ogni giorno più dati video che nei primi cinquant’anni di televisione. No, non sono questi i numeri effettivi sulla mole e sulla natura dei dati di oggi – li ho inventati – ma probabilmente non si discostano di molto da quelli veri.

Qui non inizio con la medesima litania perché ritengo che non sia poi così importante. Sì, nel mondo esiste una mole enorme di dati. Secondo uno studio, il mondo nel 2012 ha usato oltre 2,8 zettabyte di dati (2.800 miliardi di gigabyte, un numero difficilmente concepibile)¹. Sì, è più di qualsiasi cosa finora conosciuta, e nel corso del tempo aumenterà ulteriormente. Peraltro, menzionare quanta sia la mole dei dati oggi disponibili può tornare utile tutt’al più per fare sfoggio delle proprie conoscenze durante una cena, ma per le organizzazioni che devono gestire i big data e trarne un vero vantaggio il volume totale non è essenziale. Potremmo persino attenerci al vec-

¹ Gantz J., Reinsel D., “Big Data, Bigger Digital Shadows, and Biggest Growth in the Far East”, 1° dicembre 2012, www.emc.com/collateral/analyst-reports/idc-the-digitaluniverse-in-2020.pdf.

chio detto (da un altro contesto), che “le dimensioni non contano”. Si tratta infatti non di stupirsi per il volume dei dati, bensì di analizzarlo, per poi convertirlo in insight, innovazioni e business value. Lo studio sopra citato indica che solamente il cinque per mille dei 2.800 miliardi di gigabyte di dati viene sottoposto ad analisi, quale che essa sia. L’ostacolo principale all’analisi è che occorre preliminarmente applicare una struttura ai big data; la maggior parte di quei 2.800 miliardi di gigabyte non sono organizzati nel formato matriciale, per righe e colonne. Dobbiamo compiere una fatica erculeica: anzitutto strutturare i dati, quindi analizzarli ed estrarne valore. Non tutto darà frutti. Nello studio citato si stima che solamente il 25 per cento abbia un potenziale valore. Quale che sia il numero esatto, stiamo appena sfiorando il vastissimo campo delle possibilità che si stanno aprendo.

1.1. Oltre la moda dei big data

Probabilmente siete scettici sui big data e su tutto il clamore che li accompagna. Così è stato anche nel mio caso, almeno finché non ho incominciato a studiare l’argomento. Ho lavorato molto con le aziende sull’impiego dell’analytics nelle loro attività e ho scritto parecchi libri in materia, fra i quali *Competing on Analytics* (insieme a Jeanne Harris) e *Analytics at Work* (insieme a Jeanne Harris e Bob Morison). Ho collaborato con più di cento aziende su come applicare l’analytics per ottenere un vantaggio competitivo.

In un primo tempo, lo confesso, pensavo che i big data non fossero altro che una riedizione dell’analytics tradizionale. L’espressione ha cominciato a circolare nell’ultimo trimestre del 2010 e non la si incontrava spesso al di fuori della Silicon Valley. Ritenevo quindi che fosse semplicemente una delle tante nuove mode lanciate dall’azione combinata di venditori, consulenti e tecnologi. Presi brevemente in considerazione addirittura l’eventualità di sostituire nei miei vecchi libri la parola “analytics” con la nuova espressione “big data”, allo scopo di ottenere senza fatica parecchi nuovi libri da pubblicare: quello sì, un bell’imbroglio!

Dopo avere iniziato ad approfondire l’argomento, nel 2011, mi sono reso conto che il mio scetticismo era infondato. Ho intrapreso parecchi studi sistematici, fra i quali uno sui data scientist e sul fattore umano nei big data, uno sui big data nelle grandi aziende, uno sui big data nel settore dei viaggi e uno sul processo di discovery applicato ai big data². Per que-

² Studio sui data scientist: T. H. Davenport, *The Human Side of Big Data e High-Performance Analytics* (finanziato da SAS ed EMC), www.sas.com/reg/gen/corp/2154478; studio sui big data nelle grandi aziende: T. H. Davenport, J. Dyché, *Big Data in Big Compa-*

sti progetti ho effettuato più di cento colloqui in imprese start-up che si occupano di big data, in imprese già affermate nell'online e in grandi aziende appartenenti a settori tradizionali. In queste ultime non era raro che i manager condividessero il mio scetticismo e la mia diffidenza nei confronti di questa nuova moda. In diversi casi essi ritenevano di trattare già da parecchi anni i big data (almeno se con questa espressione si intende principalmente una grande mole di dati) e sostenevano che in materia non c'era nulla di nuovo, a ben vedere. Nel corso dei colloqui però molti di loro ammettevano che l'assenza di struttura in gran parte dei dati disponibili oggi dà luogo a nuove difficoltà, come anche a nuove opportunità di business.

Alla fine della mia indagine ho concluso che vi sono differenze sostanziali fra l'analytics tradizionale e i big data (si veda la tab. 1.1), anche se questo non sempre emerge dalla lettura di altri libri e articoli sull'argomento, spesso alquanto confusi. Spiegherò queste differenze, e alcune importanti analogie, più avanti in questo capitolo e nel resto del libro. La mia tesi è comunque che i big data, nonostante le mie riserve sull'appropriatezza dell'espressione rispetto al fenomeno, non sono affatto una moda passeggera e avranno anzi una notevole importanza per numerose organizzazioni. Spero di persuadere anche voi. Non rinunciate tuttavia subito al vostro scetticismo.

Tab. 1.1 – Big data e analytics tradizionale

	Big data	Analytics tradizionale
Tipo di dati	Non strutturati	Ordinati per righe e colonne
Volume dei dati	Da cento terabyte ai petabyte	Decine di terabyte, o meno
Flusso dei dati	Flusso costante di dati	Pool statico di dati
Metodi di analisi	Machine learning	Per ipotesi
Scopo principale	Prodotti data-based	Servizi, supporto alle decisioni

In fin dei conti, se vi convincerete che i big data sono importanti per voi e per la vostra organizzazione, dovrete anche agire di conseguenza. Dovrete decidere quali aspetti dei big data siano più adatti alla vostra attività e iniziare a occuparne. Potrete anche avere bisogno di assumere, o iniziare

nies Research Report (finanziato da SAS), www.sas.com/reg/gen/corp/2266746; studio sui big data nel settore turistico: T. H. Davenport, *At the Big Data Crossroads: Turning Toward a Smarter Travel Experience* (finanziato da Amadeus), www.amadeus.com/bigdata; studio sul data discovery (finanziato da Teradata Aster) ancora in corso all'epoca della pubblicazione di questo libro.

a formare, il tipo di personale che sia in grado di rendere funzionali i big data. E dovrete magari anche cambiare la vostra architettura tecnologica. Ritengo che questo libro avrà ottenuto il suo scopo se inizierete a fare queste cose. Solo però se saranno adatte al vostro caso e comunque solo dopo che ve ne sarete pienamente persuasi.

Il mio intento è quindi non solo di presentarvi i big data, ma anche di aiutarvi a prendere in merito decisioni efficaci. Vi dirò che cosa mi ha colpito dei big data, indicando però anche quali aspetti vengono esagerati. Vi dirò quali settori e quali organizzazioni ritengo che ne subiranno profonde trasformazioni, ma anche quali ne resteranno relativamente immuni, almeno per un certo tempo. E se deciderete di procedere con i big data, vi mostrerò i modi più sensati e più economici in cui potrete utilizzarli.

Non intendo abusare del vostro tempo e della vostra attenzione, tuttavia spero che siate disposti a mobilitare risorse intorno a un progetto sui big data. Potrà essere un semplice progetto esplorativo, o un progetto che inizia a costruire alcune competenze, ma sono convinto che nella maggior parte dei casi le organizzazioni debbano muoversi in questa direzione. Quanto meno, il vostro team di management dovrebbe interrogarsi su come i big data possono adattarsi al vostro business.

Probabilmente è meglio incominciare da quello che mi immagino di voi, in quanto miei lettori. Poi spiegherò alcune delle mie riserve sui big data e sul gran vociare che si fa attorno a questo argomento.

Nel resto del capitolo vi dirò perché ritengo che i “big data” siano un “big deal”, un grande affare.

1.2. Chi siete?

Chi siete per interessarvi al valore dei big data e alle loro implicazioni per il mondo del business? Mi sono fatto alcune idee a questo riguardo (anche se forse avete trovato il libro in una sala d’aspetto, o l’avete selezionato involontariamente per un download su Kindle!).

I lettori dei libri che ho già pubblicato sull’analytics tradizionale erano normalmente uomini d’affari, brillanti e ambiziosi, che credevano nella capacità dei dati e della tecnologia di cambiare le regole del gioco. Ho motivo di ritenere che anche voi siate così.

Più precisamente, tuttavia, ho constatato che gli executive maggiormente interessati ai big data provengono da funzioni data-intensive come il marketing, la logistica, e sempre più frequentemente anche dalla finanza e dalle risorse umane. Anche gli executive appartenenti a funzioni or-

ganizzative di supporto ai big data, come l'information technology (IT), dimostrano spesso un notevole interesse a saperne di più. Gli executive maggiormente interessati lavorano in settori che sono già fortemente orientati ai dati, come quello delle aziende che operano esclusivamente online, o in quelli in cui i dati hanno un grosso potenziale di cambiamento – in linea di principio, qualsiasi settore, ma specialmente quelli dove sono disponibili grandi moli di dati sui consumatori – per esempio il retail, il turismo, i trasporti, le telecomunicazioni, i media e l'entertainment, i servizi finanziari.

Infine potreste anche essere studenti che si preparano a una carriera – quanto meno, a un paio di impieghi – nei big data. In tal caso, rallegratevi per la vostra perspicacia. Questa è una disciplina che probabilmente vivrà un vero e proprio boom per parecchi anni.

Se siete come la maggior parte dei manager e dei professionisti che incontro, sarete certamente a conoscenza del fatto che i dati sono andati espandendosi a un ritmo impressionante. Potreste però avere qualche dubbio a proposito di che cosa distingua veramente i big data dal data management e dall'analytics più tradizionali. Non siete i soli. Molti dei quasi mille lettori della *Harvard Business Review* interpellati nel corso di un'indagine condotta dalla rivista nel 2013 hanno risposto di avere familiarità con il concetto di “big data”. Solo il 28 per cento ha però dichiarato di ritenere la propria organizzazione pronta a “utilizzare i big data per migliorare le decisioni commerciali o per creare nuove opportunità di business”. E solamente per il 23 per cento la propria organizzazione disponeva di una strategia sui big data. A riconoscersi nell'affermazione “la mia organizzazione ha considerato l'impatto dei big data sulle funzioni primarie per il suo business” era poi solo il 6 per cento del campione intervistato. E una percentuale ancor più esigua (il 3,5 per cento) si riconosceva in quest'altra affermazione: “la mia organizzazione sa come applicare i big data al proprio business”. Ciò probabilmente spiega il motivo per cui vi siete rivolti a questo libro: aiutare la vostra organizzazione ad affrontare tali questioni, e magari promuovere così facendo la vostra carriera. Mi piace pensare che siate capitati nel posto giusto!

1.3. Decostruire l'espressione “big data”

Come anche molti degli executive con i quali ho parlato di questo fenomeno, apprezzo tutto del concetto di “big data” a eccezione del nome. Il concetto, come ho già osservato, è rivoluzionario e promette profonde tra-

sformazioni in quasi tutti i campi di attività. Il nome invece è problematico, per diverse ragioni. Anzitutto, perché la grande mole, il “big”, non è che un aspetto di ciò che caratterizza nuove forme di dati, e per diverse organizzazioni non è neppure l’aspetto più importante. Un’indagine del 2012 di NewVantage Partners su più di cinquanta executive in grandi organizzazioni ha mostrato che per le grandi aziende trattare l’assenza di struttura nei dati è più cruciale che trattarne semplicemente la grande mole. Il 30 per cento di coloro che hanno risposto all’indagine affermava infatti che la principale ragione del loro interesse per i big data fosse quella di “analizzare dati provenienti da fonti diverse”, mentre per un altro 22 per cento essa dipendeva dalla necessità di “analizzare nuovi tipi di dati”. Per il 12 per cento infine la ragione che li spingeva a considerare i big data era: “analizzare dati di tipo streaming”. Solamente nel 28 per cento dei casi coloro che hanno risposto si occupavano di analizzare dataset da un terabyte o più, e fra questi erano in tanti (il 13 per cento) a orientarsi verso dataset di grandezza compresa fra uno e cento terabyte, un volume non molto grande, secondo gli standard dei big data³. Vi sono poi ulteriori problemi relativi all’espressione “big data”. L’attributo “big” è ovviamente relativo: ciò che oggi è grande, domani non lo sarà poi così tanto e, come indica l’indagine già menzionata, ciò che è grande per un’organizzazione può essere piccolo per un’altra. Io generalmente associo l’aggettivo “big” a un volume di dati di almeno un decimo di petabyte, tuttavia l’unico aspetto per il quale “le dimensioni contano” nei dati è l’entità dell’hardware che occorre acquistare allo scopo di archivarli e di elaborarli.

Alcuni hanno definito i big data sulla base delle *tre V* (volume, varietà, velocità); altri hanno aggiunto ulteriori V (veracità, valore e, forse, la prossima sarà venalità) ma anche questo è problematico. Certamente si tratta di caratteristiche importanti, ma che dire se si è in presenza solamente di una delle V? Oppure di due? Si avranno big data solamente per un terzo, o per due quinti? Un ulteriore problema è che troppe persone – soprattutto i venditori – stanno già usando l’espressione “big data” per indicare un impiego qualsiasi dell’analytics tradizionale, quando non addirittura del reporting e della business intelligence convenzionale.

È cosa nota che una parte dei venditori e dei consulenti faccia spesso ricorso a qualsiasi parola che sia nuova e alla moda per descrivere la solita offerta. Ciò, per quanto riguarda i big data, è già accaduto in moltissime occasioni.

³ NewVantage Partners, *Big Data Executive Survey: Themes and Trends*, 2012, <http://newvantage.com/data-management>. Segnalo che sono consulente di NewVantage Partners.

Siate perciò cauti, quando leggete libri, articoli o comunicati pubblicitari sui big data; se si riferiscono al reporting, al decision-making data-driven o all’analytics convenzionale, avete a che fare con idee utili e apprezzabili, ma non particolarmente nuove.

A causa di questi problemi definitori, io (insieme ad altri esperti che ho consultato) prevedo una vita relativamente breve per questa infelice espressione. Essa è stata adottata dai media e dalle start-up, ma ho già constatato una certa riluttanza a farla propria da parte di chi si occupa di tecnologia nelle grandi organizzazioni, particolarmente quelle che gestiscono da molti anni grandi volumi di dati, come le banche e le aziende di logistica e trasporti. Come illustrerò con maggiore dettaglio nel capitolo 8, queste persone sentono che l’attuale generazione di nuove fonti e nuovi tipi di dati non è che un ulteriore anello che si aggiunge a una lunga catena di generazioni precedenti. Naturalmente questo non vuol dire che il fenomeno noto come big data si esaurirà. Per descrivere l’ampia gamma dei nuovi e massivi tipi di dati che sono comparsi nell’ultimo decennio, o giù di lì, non conosco un’altra denominazione generica che sia migliore di “big data”.

Poiché però l’espressione è così imprecisa, le organizzazioni devono imparare a decostruirla in modo da affinare le loro strategie, segnalando agli stakeholder ciò che davvero sono interessate a fare con questi nuovi tipi di dati e quali sono per loro quelli di maggiore importanza. Esistono naturalmente numerose varianti di big data fra le quali scegliere (come si vede nella tab. 1.2, il ventaglio di opzioni è veramente ampio). Potete iniziare scegliendone una da ciascuna colonna.

Tab. 1.2 – Possibili varianti di big data

Stile dei dati	Fonte dei dati	Settore interessato	Funzione interessata
Grande volume	Online	Servizi finanziari	Marketing
Non strutturati	Video	Sanità	Logistica
Flusso continuo	Sensori	Manifattura	Risorse umane
Formati multipli	Genomica	Turismo/trasporti	Finanza

In altre parole, invece di dire “stiamo avviando un’iniziativa di big data” è probabilmente più costruttivo dire “intendiamo analizzare i dati video dai nostri bancomat e dai nostri sportelli di agenzia allo scopo di capire meglio le relazioni con la clientela”. Oppure, nel settore sanitario, si può decidere di “combinare i dati sanitari in formato elettronico e i dati genomici per applicare ai pazienti trattamenti personalizzati”. Oltre a fare

chiarezza sulle intenzioni e le strategie, questo approccio consente di evitare interminabili discussioni sul tipo di dati, se “big” o “non big” (in effetti, poche organizzazioni ammettono di avere a che fare con “small data”, benché non vi sia assolutamente nulla di male: mi hanno insegnato che un’espressione è veramente utile solamente se anche il suo opposto è considerato valido).

Naturalmente, potete usare un’espressione alla moda perché volete ottenerne un qualche vantaggio. Se la vostra organizzazione è del tipo che risponde solamente a temi manageriali nuovi di zecca – e se quello dei big data sarà ancora un tema relativamente nuovo di zecca quando leggerete questo libro – avviate senz’altro un progetto sui big data. In altre parole, se l’espressione “big data” serve a mobilitare energie ed entusiasmo nella vostra impresa, usatela pure. Siate però lesti a convertirvi alla prossima espressione, quando la moda cambierà; IBM ha già un centro di ricerca sui “massive data”, non tarderemo a leggere di “gargantuan data” o “humongous data”! In questo libro continuerò a usare l’espressione “big data” perché al momento non ce ne è un’altra utile a meglio descrivere l’insieme di attributi che essa denota. Sono però convinto che per estrarre valore autentico da questa risorsa dal nome così inadeguato sia necessario un attento approfondimento.

1.4. I big data dureranno?

Possiamo forse ammettere che i big data abbiano un nome infelice. Ma si tratta di qualcosa destinato a durare nel tempo oppure di una moda passeggera? Che siano lo hula-hoop o la pet rock⁴ del mondo dell’information management? Se sì, i business leader dovrebbero tranquillamente ignorarli. Se invece essi diventeranno un aspetto fondamentale per le aziende, i manager che li ignoreranno lo faranno a loro rischio e pericolo.

Vi sono certamente alcuni tratti tipici delle mode passeggiere nell’idea di big data, tanto più che la concezione alla quale essa si ispira non è particolarmente nuova.

L’idea di analizzare i dati per capire che cosa stia succedendo alle nostre attività commerciali ci accompagna da un bel po’ di tempo (nelle corporation almeno dal 1954, quando UPS formò un gruppo dedicato all’analisi dei dati). Perché mai dunque dovremmo inventarci continuamente

⁴ Le “pietre domestiche”, ciottoli grigi commercializzati in una scatola di cartone provvista di fori per l’areazione e accompagnata da un comico libretto di istruzioni con indicato il modo in cui prendersi cura delle proprie pietre, furono una moda di successo in America negli anni Settanta [N.d.R.].

dei nuovi nomi per esprimerla? L'attività di estrarre senso dai dati è stata negli anni chiamata "ausilio alle decisioni" (*decision support*), "supporto direzionale" (*executive support*), "Olap" (*online analytical processing*), "business intelligence", "analytics" (si veda la tab. 1.3)⁵. Vi sono certamente alcune novità in ciascuna delle diverse denominazioni che si sono susseguite, eppure non sono così certo che vi sia stata, nelle cose, un'evoluzione così spinta da meritare sei generazioni di definizioni.

Tab. 1.3 – Terminologia utilizzata per descrivere l'uso e l'analisi dei dati

Termine	Periodo	Significato specifico
Ausilio alle decisioni (Decision support)	1970-1985	Impiego dell'analisi dei dati in ausilio al decision-making
Supporto direzionale (Executive support)	1980-1990	Impiego dell'analisi dei dati in ausilio alle decisioni del top management
Online analytical processing (Olap)	1990-2000	Software per l'analisi di data table multidimensionali
Business intelligence	1989-2005	Tool di ausilio a decisioni data-driven, con enfasi sul reporting
Analytics	2005-2010	Accento sull'analisi matematica e statistica in ausilio alle decisioni
Big data	dal 2010	Accento su dati massivi, non strutturati, a elevata mobilità

Ciò che può almeno in parte giustificare un nuovo nome per i big data è la presenza di grandi volumi di dati di nuovo tipo. Secondo una stima, vengono generati ogni giorno nel mondo 2,5 quintilioni (ovvero 2,5 seguito da diciotto zeri) di byte⁶. I dati di nuovo tipo, non strutturati, sono meritevoli di una nuova denominazione e di un nuovo trattamento, come ho già osservato. Certi tipi di dati, come il testo e la voce, ci accompagnano da lunghissimo tempo, ma sono la loro mole su Internet e in altre forme digitali, nonché le nuove tecnologie che consentono di analizzarli, a segnare il passaggio a una nuova epoca.

Del tutto nuovi sono invece i dati dei social media, dalle pagine di Facebook e Pinterest fino ai tweet. Non so se queste nuove forme di dati resteranno tutte in circolazione. È probabile che le funzioni caratteristiche di questi tool vengano incorporate in applicazioni di più ampio spettro. I dati

⁵ Power D., "A Brief History of Decision Support Systems". Per maggiori dettagli sulla terminologia dei primi tempi si veda: <http://dssresources.com/history/dsshistory.html>.

⁶ La stima dei 2,5 quintilioni di byte viene da IBM, "What Is Big Data? Bringing Big Data to the Enterprise", www.ibm.com.