

BIG DATA

The background of the cover features a complex network diagram. It consists of numerous circular nodes of varying shades of blue and black, interconnected by a web of solid and dashed lines. The nodes are scattered across the entire page, creating a sense of a vast, interconnected data network.

'Data and analytics power everything that we do. This book is the go-to-guide on data for 2015.'

— Henrik von Scheel,
Advisory Board Member
at Google

USING
SMART BIG DATA
ANALYTICS AND
METRICS TO MAKE
BETTER DECISIONS
AND IMPROVE
PERFORMANCE

BERNARD
MARR

‘While Bernard is clearly ahead of the pack on Big Data, the brilliance of his book is in its timing. Just when we were starting to feel let down by the over-hyped promise, he brings Big Data into sharp focus and spells out his tried and trusted (“SMART”) approach to help us tackle the greatest opportunity (or threat) of our generation. No more excuses!’

Richard Phillips, Director of Analytics, Barclaycard

‘What a thought-provoking and enjoyable read! The powerful, yet simple, SMART model will allow anyone to take advantage of Big Data in so many different ways, from improving profitability and customer retention to winning sports games! Each chapter will trigger fresh ideas and identify new opportunities to better leverage data in your company.’

Marcus Barlow, Operations Director at American Express

‘Data and analytics power everything that we do. This book is the go-to-guide on data for 2015. A brilliant piece of work!’

Henrik von Scheel, Advisory Board Member at Google, EMEA and Gazprom, CEO of LEADing Practice

‘Bernard Marr is a master at synthesizing a complex set of topics into salient points that practitioners need to know. In his newest book, Bernard has boiled down the Big Data ocean into a simple and practical SMART methodology that will help organizations extract real value from a dizzying array of data, tools, and technologies.’

Wayne Eckerson, Principal Consultant, Eckerson Group

‘Bernard Marr has done it again – taken a complex subject and broken it down into simple pieces so that business leaders can devise practical strategies for exploiting the opportunities presented by Big Data. This book is a must-read for anyone trying to understand and leverage Big Data.’

Dave Kellogg, CEO, HostAnalytics

‘This is a SMART book by a SMART author. Bernard Marr goes beyond the hype of Big Data, providing real-life case studies and action points for the manager looking for the competitive edge.’

Lars Rinnan, CEO, Nextbridge

‘This book will help you unravel the mystery of Big Data. It simply lifts any confusion caused by buzzwords and technical terms that are thrown about when people talk about Big Data. The book provides many examples of organizations making sense of a variety of data to achieve real business impact. The book’s “SMART” approach will help you avoid the common and expensive mistake of gathering a mountain of data with no notion of what to do with it.’

Robert Stackowiak, Vice President of Information Architecture and Big Data, Oracle

‘Being smart, Bernard Marr has created an enjoyable book that describes the world of Big Data and analytics and how this will completely change our business world.’

Professor Kai Mertins, President, Interop VLab

‘Is Big Data a buzzword or does it have practical applications in business? Bernard Marr goes beyond the hype of Big Data to provide business people with a smart solution to understand where we are, where we are trying to get to and what data and tools we can employ to help us get there.’

Roberto Croci, Manager, Google Analytics (SEEMEA)

‘I would recommend this book to anyone looking to put an efficient data-driven strategy in place. The SMART methodology is a simple way of bringing together the key concepts of an analytics strategy.’

Benjamin Mercier, Vice President, Senior Digital Analytics Manager at Barclays

‘Bernard Marr’s *Big Data* provides a lucid, practical guide for managers seeking to leverage the explosion in data and data analysis for productivity. In a convincing fashion, Marr moves beyond traditional understandings of the emergence of Big Data in business practices as a technical issue and construes it as a managerial issue rooted in both leadership and strategy. The message is that data is meaningless without the ability to understand it or to communicate what that understanding is.’

Robert M. Mauro, PhD, Director and Founder, Global Leadership Institute, Boston College

'In a world where 90% of all data has been created in the last two years, Big Data and analytics offer massive opportunities for consumers, corporates and governments. Bernard's book *Big Data* should be essential reading for any manager because it provides you with a pragmatic guide to realizing the real value of the fourth production factor: data! It is a very practical book that will take you beyond the hype of Big Data. Using Bernard's SMART model on data analytics will help you dramatically improve performance through data-driven decision making!'

Berry Diepeveen, Partner, Enterprise Intelligence & Analytics, Ernst & Young

'This is a practical, insightful, no-nonsense book on Big Data that will provide a 360-degree view of how Big Data impacts our life and business. Bernard is an awesome simplifier and thought leader who presents the complex subject of Big Data in a way that everyone can understand.'

Nandhini Sampath, Sr. Manager of Business Transformation & Analytics, Cisco Systems

'Bernard leverages his years of experience to provide a practical roadmap for quickly realizing real returns on your Big Data journey.'

Kurt J. Bilafer, Vice President, ClearStory Data

'You can cut through the hype associated with the latest management fad, which is Big Data, with the help of Bernard Marr who reminds us that "the real value is not in the large volumes of data but what we can now do with it". And "The reality is that most businesses are already data rich, but insight poor". The recommendation is, focus on the SMART Data, not the Big Data. In doing so, regardless of size or budget, you can harness the advantages and become a SMART Business.'

Paul Barnett, Founder & CEO, Strategic Management Forum

'In the midst of an unstructured world of novel possibilities, I finally found a consistent framework that will help make Big Data a reality in our business strategy execution. It's not about technology, but how to transform our business to cope with the new century of opportunities.'

Pedro Pereira, Head of SAP Big Data, SAP

‘Unlike so many other books about Big Data, this one is focused on the business value of Big Data. This practical book provides a step-by-step approach and countless real-world examples of how to turn data into smart insights that can transform not only businesses but also the ways we deal with social, political and health issues. Bernard has written a must-read book for anyone who wants to understand the data and analytics revolution.’

**Lucile Hyon-Le Gourrièrec, Big Data & Analytics Marketing Leader,
IBM France**

‘Many boards and executives across the public and private sector stuck their heads in the sand when it came to IT – at great shareholder and taxpayer cost. Don’t make the same mistake again with Big Data! There is no doubt that Big Data and analytics, driven by technology and business model innovation, is profoundly reshaping and disrupting industries, economies and society at an unprecedented rate. Bernard Marr provides a thought provoking and gripping insight into the power of Big Data at the heart of this smart revolution. Big Data is compelling reading.’

**Marie Johnson, Chief Digital Officer, Centre for Digital Business
and formerly Microsoft’s Worldwide Director of Public Services and
eGovernment**

BIG DATA

**USING SMART BIG DATA,
ANALYTICS AND METRICS TO
MAKE BETTER DECISIONS
AND IMPROVE
PERFORMANCE**

BERNARD MARR

WILEY

This edition first published 2015

© 2015 John Wiley & Sons Ltd

Registered office

John Wiley & Sons Ltd, The Atrium, Southern Gate, Chichester, West Sussex, PO19 8SQ, United Kingdom

For details of our global editorial offices, for customer services and for information about how to apply for permission to reuse the copyright material in this book please see our website at www.wiley.com.

All rights reserved. No part of this publication may be reproduced, stored in a retrieval system, or transmitted, in any form or by any means, electronic, mechanical, photocopying, recording or otherwise, except as permitted by the UK Copyright, Designs and Patents Act 1988, without the prior permission of the publisher.

Wiley publishes in a variety of print and electronic formats and by print-on-demand. Some material included with standard print versions of this book may not be included in e-books or in print-on-demand. If this book refers to media such as a CD or DVD that is not included in the version you purchased, you may download this material at <http://booksupport.wiley.com>. For more information about Wiley products, visit www.wiley.com.

Designations used by companies to distinguish their products are often claimed as trademarks. All brand names and product names used in this book and on its cover are trade names, service marks, trademark or registered trademarks of their respective owners. The publisher and the book are not associated with any product or vendor mentioned in this book. None of the companies referenced within the book have endorsed the book.

Limit of Liability/Disclaimer of Warranty: While the publisher and author have used their best efforts in preparing this book, they make no representations or warranties with respect to the accuracy or completeness of the contents of this book and specifically disclaim any implied warranties of merchantability or fitness for a particular purpose. It is sold on the understanding that the publisher is not engaged in rendering professional services and neither the publisher nor the author shall be liable for damages arising herefrom. If professional advice or other expert assistance is required, the services of a competent professional should be sought.

Library of Congress Cataloging-in-Publication Data

Marr, Bernard.

Big data : using smart big data, analytics and metrics to make better decisions and improve performance / Bernard Marr.

pages cm

Includes index.

ISBN 978-1-118-96583-2 (pbk.)

1. Information technology—Management. 2. Big data. 3. Management—Statistical methods.

4. Decision making—Statistical methods. I. Title.

HD30.2.M3744 2015

658.4'0380285574—dc23

2014040562

A catalogue record for this book is available from the British Library.

ISBN 978-1-118-96583-2 (pbk)

ISBN 978-1-118-96582-5 (ebk) ISBN 978-1-118-96578-8 (ebk)

Cover Design: Wiley

Cover Image: iStockphoto.com/marigold_88

Set in 10.5/14pt MinionPro Light by Aptara Inc., New Delhi, India

Printed in Great Britain by TJ International Ltd, Padstow, Cornwall, UK

To the most important people in my life:
My wife Claire and our three children Sophia,
James and Oliver;
as well as my brother Marc, Julie and Alan, all my wonderful
friends, and in memory of my parents.

CONTENTS

	Introduction: Welcome to a Smarter World	1
1	Smarter Business	9
2	S = START WITH STRATEGY	23
3	M = MEASURE METRICS AND DATA	57
4	A = APPLY ANALYTICS	105
5	R = REPORT RESULTS	155
6	T = TRANSFORM BUSINESS	199
	Conclusion	231
	<i>About the Author</i>	236
	<i>Acknowledgements</i>	238
	<i>Index</i>	239

INTRODUCTION: WELCOME TO A SMARTER WORLD

The world is getting smarter.

This evolution can be seen everywhere and no industry or sector is immune. Consider an industry as old and well established as fishing, for example. Although human beings have been fishing since the beginning it wasn't until the 16th century that fisherman had boats capable of going to sea. This advance radically changed the fortunes of fishing and made large, profitable catches possible for the first time. The ships would set out for the fishing grounds using little more than a compass, a sextant and some 'inside knowledge' passed down through the generations of fishing families. If they were sailing at night they would use celestial navigation techniques and plot a course by the stars in order to arrive in the right vicinity. When the fisherman arrived at the fishing grounds they would cast their nets and hope for the best.

By the late 19th century fishing had been commercialized. Small fishing boats gave way to massive trawlers with on-board processing capabilities, the discovery of longitude and latitude made

BIG DATA

navigation considerably easier and in the last few decades technology has transformed fishing from an art to a science. Modern fishing boats are technology rich, using high tech navigation systems and GPS. Often small sensors are attached to the fish to track where the shoals are at any given time and sonar is used to pinpoint the density of the shoal and where and when to cast the nets. Modern fisherman know where the fish are, they know where they will be tomorrow and when to cast their nets for the best possible catch of their target fish. Fishing has evolved to become smarter. And it is just one example. Today the world is smarter in everything from sport's performance to healthcare in the home. Even love and parenting is becoming smarter!

Smarter sport

Smart technology is now widely used in sport to find and recruit talent as well as monitor and improve performance – both for the amateur and the professional. It's now possible to get a basketball with over 200 built-in sensors that provide player and coaches with detailed feedback on performance. In tennis a system called SlamTracker can record a player's performance providing real-time statistics and comprehensive match analytics. If you've ever watched rugby (union or league) you may have wondered what the bump is between the players' shoulder blades – it's a GPS tracking system that allows the coaching staff to assess performance in real time. The device will measure the players' average speed, whether the player is performing above or below their normal levels, and heart rate, to identify potential problems before they occur. All of which can help coaches avoid injury and assist in making appropriate substitution decisions. Similar technology exists in the English Premier League and is used by many Olympic sports such as cycling.

INTRODUCTION

But the technology is not just for the professionals. There are many wearable devices that can monitor health and well-being on the go. For example I wear an ‘Up’ fitness band that tells me how many steps I have taken each day, how many calories I’ve burned and how well I’ve slept each night. It is synced to my bathroom scales so that if I put on weight it will tell me and prompt me to increase my activity or decrease my food intake.

Smarter health

Healthcare is also becoming smarter and it’s set to revolutionize our lives.

Professor Larry Smarr, one of the most influential computer scientists in the United States and the most monitored man on the planet was able to self-diagnose Crohn’s disease – long before any symptoms emerged and early enough to be able to effectively manage the condition. Smarr states:

‘In a world in which you can see what you are doing to yourself as you go along the hope is that people will take more personal responsibility for themselves, in keeping themselves healthy. So it’s almost like we are at day zero of a whole new world of medicine, and what will come out the other end is a far healthier society that’s focused on wellness rather than fixing sickness when it’s way too late.’¹

This ability to monitor our own health heralds a new and exciting frontier of preventative medicine based on data.

¹BBC Two (2013) Horizon Monitor Me narrated by Dr Kevin Fong (2013).

BIG DATA

We have long understood that in theory prevention is better than cure but the collaboration of technology and health is turning that insight into practice. This year close to 42 million wearable wireless sports, fitness and wellness devices are expected to ship worldwide. According to ABI Research, 'Over the next five years spending on bringing these wearable wireless consumer activity device collected data will grow to a \$52 million market by 2019'. Cloud services such as Ginger.io already allow care providers to monitor their patients through sensor-based applications on their smart phones.² And Proteus manufactures an 'ingestible' scanner the size of a grain of sand, which can be used to track when and how patients are taking their medication. This gives providers information about 'compliance rates' – how often patients follow their doctor's orders – and can even alert a family member to remind them.

But it's not just the ability for us to monitor and manage our own health better; Big Data, analytics and the smart revolution are changing healthcare right now with innovations such as state-of-the-art brain injury scanners, premature baby units and cancer detection and diagnosis systems. The possibilities are endless.

Smarter homes

Everything is also getting smarter at home. From the cars we drive to and from home, to the heating systems, gadgets, appliances and even the carpet!

²Palmer, S., White, E., Romanski, P., Benedict, K. and Gardner, D. (2014) Integrating Consumer Wearable Health Devices Will Drive Healthcare Big Data Adoption, Says ABI Research. <http://bigdata.ulitzer.com/node/3058905>

INTRODUCTION

The evolution from basic to smart is especially noticeable with cars. Initially the Model T Ford was black, stick shift, a few buttons and no seat belt. Today we have cars with dashboards that resemble an aircraft cockpit, with cameras and sensors for easy parking, alerting the driver if he or she gets too close to the kerb or another car. Some cars will parallel park themselves and brake automatically. Others will sync with traffic information and redirect you to a better route to avoid traffic black spots or an accident. Sensors on the engine will monitor how well you are driving, which will in turn potentially lower (or raise) your insurance and dynamically adjust your service intervals based on your driving style.

There are smart thermostats that monitor the home and only heat the areas that are being used. The temperature of your home can be changed while you are still at work so that when you arrive on a winter's evening the house is cosy. This ability to monitor and dynamically alter temperature can save energy and money. Obviously solving the energy crisis is not just about finding new energy sources such as wind and solar but also about saving the energy we have and using it more efficiently.

Smart TVs use face recognition to make sure your children don't ever watch anything unsuitable for their age and smart carpets can alert you should your elderly parent not make their usual morning coffee.

Considering all the toys, gadgets and smart appliances there are now more machines connected to the Internet than people. And all those smart things are gathering data and communicating with each other.

Smarter love

Even something as personal and magical as falling in love is getting smarter. Everyone hopes to find their soul mate and yet, for many the search is far from straightforward. Online dating site eHarmony matches people based on twenty-nine different variables such as personality traits, behaviours, beliefs, values and social skills. Each person who joins eHarmony completes a comprehensive profile questionnaire, which provides the data for the analytics model to find potential matches.

US digital specialist, Amy Webb, even took the online data algorithms one step further. After one particularly terrible first date where her 'Prince Charming' ordered the most expensive items from the menu, enjoyed them and did a runner after excusing himself for the bathroom, Webb created her own personal scoring system based on what was important to her in a potential life partner. In addition she analysed other profiles to see what attracted attention, tested changes to her own profile to see what made a difference to the number and quality of enquires and would only agree to go on a date with someone if he scored above a certain number. And it worked... Amy Webb is now happily married and the couple have a daughter.³

Smarter parenting

The complex art of parenting is also getting smarter. To identify and reduce potential pre- and postnatal risks, many babies around

³http://www.ted.com/talks/amy_webb_how_i_hacked_online_dating.html

INTRODUCTION

the world are being constantly monitored across a myriad of metrics and data points including heart rate and respiration. These vital measures are able to predict infections 24 hours before the baby shows any visible symptoms and can allow for early, often life-saving intervention.

Once your baby has arrived safely he or she can also sleep on a mattress full of sensors that monitor breathing patterns and heart rate and alerts parents if anything is wrong. Just imagine how many tragic cot deaths could be avoided with this smart technology. We can even buy digital diapers which will send a tweet to our smart phone when our baby needs changing! Obviously a good parent doesn't really need a tweet to tell them this information but the latest generation of these diapers automatically analyses the urine and alerts the parent of an increased sodium level, possible dehydration, as well as the onset of any infections – and all this even before any physical symptoms appear.

The marriage of data and technology is radically changing our world and making it smarter. And business must become smarter too.

Going back to the fishing analogy for a moment ... When fishing emerged as an industry, the competition was sufficiently low and the stocks of fish sufficiently high that the fisherman didn't need to be in an exact location to enjoy a prosperous day at sea. Their experience, equipment and the number of fish in the oceans meant they would be successful unless they hit particularly bad weather. Today, with intense competition and finite fish stocks that need to be responsibly managed, fishermen have had to evolve and become smarter. And the same is true for all businesses in all sectors.

BIG DATA

Today the really successful companies understand where their customers are and, perhaps more importantly, what they are doing and where they are going. They know what is happening as it's happening and they allow that information to guide their strategy and inform their decision-making.

Companies that won't embrace the SMART revolution will be left behind.

1

SMARTER BUSINESS

Big Data is at the heart of the smart revolution. The basic idea behind the phrase 'Big Data' is that everything we do is increasingly leaving a digital trace (or data), which we (and others) can use and analyse to become smarter. The driving forces in this brave new world are access to ever-increasing volumes of data and our ever-increasing technological capability to mine that data for commercial insights.

There is little doubt that Big Data is changing the world. It is already completely transforming the way we live, find love, cure cancer, conduct science, improve performance, run cities and countries and operate business. As a result there is a huge amount of hype and fuss over Big Data. Everyone is discussing it. It is THE hot topic discussed in every boardroom, every business publication from *The Economist* to *Fortune* to the *Harvard Business Review*. Big Data is even making its way into mainstream media.

But despite the noise around Big Data most people still don't really understand it and very few people know what to do about

BIG DATA

it. Personally, I don't like the term because it's too simplistic and potentially misleading. Granted, we are now tracking and storing data on everything so we potentially do have access to large volumes of data – hence the term Big Data. But the real value is not in the large volumes of data but what we can now do with it. It is not the amount of data that is making the difference but our ability to analyse vast and complex data sets beyond anything we could ever do before. Innovations such as cloud computing combined with improved network speed as well as creative techniques to analyse data have resulted in a new ability to turn vast amounts of complex data into value. What's more, the analysis can now be performed without the need to purchase or build large supercomputers. This means that any business, government body, or indeed anyone can now use Big Data to improve their decision-making.

Especially powerful is our ability to analyse so called 'unstructured data' (more on this in Chapter 3). Basically, unstructured data is the data we can't easily store and index in traditional formats or databases and includes email conversations, social media posts, video content, photos, voice recordings, sounds, etc. Combining this messy and complex data with other more traditional data is where a lot of the value lies. Many companies are starting to use Big Data analytics to complement their traditional data analysis in order to get richer and improved insights and make smarter decisions.

In effect what Big Data should really stand for is SMART Data and whilst I think the term Big Data will disappear in time, the increasing production and use of SMART Data is definitely here to stay.

Who is using Big Data?

The big players in the space, including Amazon, Google, Walmart, and Facebook, are already making a splash. Walmart, for example, handles more than a million customer transactions each hour and imports those into databases estimated to contain more than 2.5 petabytes of data.¹ The company is now able to combine data from a variety of sources including customers' past purchases and their mobile phone location data, Walmart internal stock control records, social media and information from external sources such as the weather, and initiate tailored sales promotions. For example, if you have bought any BBQ-related goods from Walmart, happen to be within a 3 mile radius of a Walmart store that has the BBQ cleaner in stock, and the weather is sunny, you might receive a voucher for money off a BBQ cleaner delivered to your smart phone!

In another example a client of mine, a leading telecom company, is using Big Data analytics to predict customer satisfaction and potential customer churn. Based on phone and text patterns as well as social media analytics, the company was able to classify customers into different categories. The analytics showed that people in one specific customer category were much more likely to cancel their contract and move to a competitor. This extremely useful information now helps the company closely monitor the satisfaction levels of these customers and prioritize actions that will prevent them from leaving and keep them happy.

¹SAS Whitepaper (2012) Big Data Meets Big Data Analytics: Three Key Technologies for Extracting Real-Time Business Value from the Big Data That Threatens to Overwhelm Traditional Computing Architectures.

BIG DATA

Even mid-tier cars today have about 40 microprocessors that measure performance. These electronics usually account for about one-third of the cost of a new car. Of course, all this data that is being generated, collected and analysed by the car manufacturers offer them significant competitive advantages. One car maker working with an external analytics company noticed that a sensor in the fuel tank made by a German supplier was not working well at all. The manufacturer could have told the supplier and asked them to fix it but then the improvement would have been passed on to other car manufacturers that use that supplier. So instead the manufacturer invented a software patch that fixed the issue, received a patent on the fix and sold the patent to the supplier.²

Big Data is changing the very nature of business, from manufacturing to healthcare to retail to agriculture and beyond. The rate that data is and can be collected on every conceivable activity means that there are increasing opportunities to fine-tune procedures and operations to squeeze out every last drop of efficiency.

How companies are using Big Data

Different industries have responded to the call in different ways. Retail and sales are seeking to collect as much information about their customers' lives as possible so as to fulfil their changing needs more effectively. Manufacturing are seeking to streamline operations. Equipment calibration settings can be recorded and refined, and product storage environments monitored to determine the optimum conditions that lead to minimum spoilage and waste.

²Mayer-Schonberger, V. and Cukier, K. (2013) *Big Data: A Revolution That Will Transform How We Live, Work and Think*. London: John Murray Publishers.

SMARTER BUSINESS

For global companies this can mean collecting and analysing data from plants across the world, allowing minor variances to be studied and their results understood.

In 2013, for example, pharmaceutical giants Merck used analysis to dramatically cut the amount of waste caused by variance in manufacturing environment conditions. It took three months and involved 15 billion calculations on individual production data from 5.5 million vaccine batches. This allowed them to discover the optimum conditions during the fermentation process, and should greatly increase their yield, once the FDA has approved the proposed changes to the manufacturing process.

In the automotive industry a 2014 report by the Centre for Automotive Research stated that advances made possible through advanced IT solutions and Big Data represented 'an engine of innovation'. The report highlighted the growing complexity of cars and the industry as the biggest challenge faced by automotive manufacturers.

The efficiency of every machine – and human – involved in the manufacturing process can be recorded so companies know what is working, and can make improvements where they are needed.

And in agriculture, data analysis is helping the industry meet the challenge of increasing the world's food production by 60%, as forecasters have said will be necessary by 2050 due to the growing population. Tractor and agricultural machinery manufacturer, John Deere, already fits sensors to its machinery. The data that is available to the farmers via its myjohndeere.com and Farmsight services

BIG DATA

helps them to establish optimum conditions for their crops. Plus the data is also used by John Deere to forecast demand for spare parts.

Of course, in business once a product has been grown or manufactured it needs to be sold and distributed. The petabytes of customer data, including you and me, already gathered by big retailers tells them who will want to buy what, where and when. Amazon, for example, uses its S3 system to keep track of millions of stock items across dozens of warehouses and distribution centres scattered around the globe. Operatives can track deliveries in real-time to see what is where, and where it should be going.

At the point of sale, retailers can use data to determine where stock should be displayed, which stores will sell most of which particular product and track customer movements around stores. Loyalty cards are not new but ever more sophisticated analysis of customer habits will lead to an increase with which retailers can predict what you will buy. This has advanced to the point where Amazon believes it will soon be able to predict what you will buy accurately enough to despatch it toward you before you have even bought it!

The connectivity that is now possible is also changing business. In 2014 Cisco announced a \$150 million fund for start-ups working on improving integration between the virtual and physical world. For a business, the ability to have its production, stock control, distribution and security systems all connected and talking to each other will mean greater efficiency and less waste. GE refers to this convergence of data and machinery as the 'Industrial