

■ ■ ■ ROB THOMAS & PATRICK McSHARRY

# BIG DATA REVOLUTION

WHAT FARMERS, DOCTORS AND INSURANCE AGENTS  
TEACH US ABOUT DISCOVERING BIG DATA PATTERNS



WILEY



**BIG  
Data  
Revolution**



**BIG**  
**Data**  
**Revolution**

**WHAT FARMERS, DOCTORS  
AND INSURANCE AGENTS  
TEACH US ABOUT DISCOVERING  
BIG DATA PATTERNS**

**Rob Thomas**  
**Patrick McSharry**

**WILEY**

This edition first published 2015  
© 2015 Rob Thomas and Patrick McSharry

*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](http://www.wiley.com).

The right of the author to be identified as the author of this work has been asserted in accordance with the Copyright, Designs and Patents Act 1988.

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 also publishes its books in a variety of electronic formats. Some content that appears in print may not be available in electronic books.

Designations used by companies to distinguish their products are often claimed as trademarks. All brand names and product names used in this book are trade names, service marks, trademarks or registered trademarks of their respective owners. The publisher is not associated with any product or vendor mentioned in this book. This publication is designed to provide accurate and authoritative information in regard to the subject matter covered. It is sold on the understanding that the publisher is not engaged in rendering professional services. If professional advice or other expert assistance is required, the services of a competent professional should be sought.

**Trademarks:** Wiley and the Wiley logo are trademarks or registered trademarks of John Wiley & Sons, Inc. and/or its affiliates in the United States and/or other countries, and may not be used without written permission. All other trademarks are the property of their respective owners. John Wiley & Sons, Ltd. is not associated with any product or vendor mentioned in the book.

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

ISBN 978-1-118-94371-7 (paperback); 978-1-118-94373-1 (ePDF);  
978-1-118-94372-4 (ePub)

Set in 9.5/11.5 MinionPro-Regular by SPS/TCS

Printed in U.S. by Bind Rite Robbinsville

This book is for those who are willing to lead, in any endeavor. Most of all, this book is for Kristin, Will, Abby, and Sam. And, a special thanks to my big sister.

— Rob Thomas

To my parents Agnes and Patrick, wife Emmeline, and children Isolde, Theodore and Caspian.

— Patrick McSharry

# About the Authors

**Rob Thomas** is Vice President of Product Development for Big Data and Information Management in the IBM Software Group. With over 15 years in the technology industry, Mr. Thomas has had the opportunity to consult to a variety of global businesses. He has experience in business and operational strategy, high technology, acquisitions and divestitures, manufacturing operations, and product design and development.

Mr. Thomas is currently responsible for product development and engineering for IBM's Big Data and Information Management product line. As Vice President of Business Development in IBM Software, Mr. Thomas led the acquisition of Netezza and Vivisimo, both leaders in the data era.

Mr. Thomas has extensive international experience, leading IBM's high technology services business in Asia Pacific, while living in Tokyo, Japan. Prior to that, he was a Partner in IBM's consulting business.

Mr. Thomas graduated from Vanderbilt University with a BA in Economics. He earned his Masters in Business Administration from the University of Florida. Mr. Thomas publishes regularly on his blog (<http://www.robdtomas.com>) and has an active following on Twitter (@robdtomas). He is an avid golfer, reader, and exercise enthusiast. He lives in New Canaan, Connecticut with his wife (Kristin) and three children (Will, Abby, and Sam).

Most of what he has learned in his life came from his parents, his wife, and his two sisters.

**Patrick McSharry** is a Senior Research Fellow at the Smith School of Enterprise and the Environment, Faculty Member of the Oxford Man Institute of Quantitative Finance at Oxford University, Visiting Professor at the Department of Electrical and Computer Engineering, Carnegie Mellon University, Fellow of the Royal Statistical Society and Senior Member of the IEEE. He takes a multidisciplinary approach to developing quantitative techniques for data science, decision-making, and risk management. His research focuses on big data, forecasting, predictive analytics, machine learning, and the analysis of human behavior. He has published over 90 peer-reviewed papers, participated in knowledge exchange programs and consults for national and international government agencies and the insurance, finance, energy, telecoms, environment, and healthcare sectors. Patrick received a first class honours BA in Theoretical Physics and an MSc in Engineering from Trinity College Dublin and a DPhil in Mathematics from Oxford University.



# About the Technical Editor

**Carin Anderson** is a freelance technical editor. She has edited, compiled, and written numerous grants and proposals over the last decade and a half. Carin developed a mobile application company, creating multi-user gaming platforms. She also co-founded an informational website targeting families with young children.

In her spare time, she enjoys spending time with her family and friends, running and reading.

## PUBLISHER'S ACKNOWLEDGEMENTS

Some of the people who helped bring this book to market include the following:

### *Editorial and Production*

VP Consumer and Technology Publishing Director: Michelle Leete

Associate Director–Book Content Management: Martin Tribe

Professional Technology & Strategy Director: Barry Pruett

Commissioning Editor: Ellie Scott

Development Editor: Tom Dinse

Copy Editor: Laura Miller

Technical Editor: Carin Anderson

### *Marketing*

Associate Marketing Director: Louise Breinholt

Marketing Manager: Lorna Mein

Marketing Executive: Polly Thomas

### *Composition Services*

Proofreader: Wordsmith Editorial

Indexer: Potomac Indexing, LLC

# Contents

<b>Prologue</b>	<b>1</b>
Berkeley, 1930s	1
Pattern Recognition	2
Nelson Peltz	3
Committing to One Percent	5
The Big Data Revolution	6
<b>Introduction</b>	<b>7</b>
Storytelling	7
Objective	7
Outline	8
Part I “The Revolution Starts Now: 9 Industries Transforming with Data”	8
Part II “Learning from Patterns in Big Data”	11
Part III “Leading the Revolution”	11
Storytelling (Continued)	13
<b>PART I: THE REVOLUTION STARTS NOW: 9 INDUSTRIES TRANSFORMING WITH DATA</b>	<b>15</b>
<b>Chapter 1: Transforming Farms with Data</b>	<b>17</b>
California, 2013	17
Brief History of Farming	18
The Data Era	19
Potato Farming	20
Precision Farming	21
Capturing Farm Data	22
Deere & Company Versus Monsanto	24
Integrated Farming Systems	25
Data Prevails	26
The Climate Corporation	26
Growsafe Systems	27
Farm of the Future	27
California, 2013 (Continued)	29

<b>Chapter 2: Why Doctors Will Have Math Degrees</b>	<b>31</b>
United States, 2014	31
The History of Medical Education	32
Scientific Method	32
Rise of Specialists	33
We Have a Problem	34
Ben Goldacre	35
Vinod Khosla	35
The Data Era	36
Collecting Data	36
Telemedicine	38
Innovating with Data	40
Implications of a Data-Driven Medical World	42
The Future of Medical School	42
A Typical Medical School	42
A Medical School for the Data Era	43
United States, 2030	44
<b>Chapter 3: Revolutionizing Insurance: Why Actuaries Will Become Data Scientists</b>	<b>45</b>
Middle of Somewhere, 2012	45
Short History of Property & Casualty	
Insurance and Underwriting	46
Actuarial Science In Insurance	47
Pensions, Insurance, Leases	49
Compound Interest	50
Probability	50
Mortality Data	50
Modern-Day Insurance	51
Eight Weeks to Eight Days	51
Online Policies	52
The Data Era	52
Dynamic Risk Management	52
Catastrophe Risk	54
Open Access Modeling	55
Opportunities	56
Middle of Somewhere, 2012 (Continued)	58
<b>Chapter 4: Personalizing Retail and Fashion</b>	<b>59</b>
Karolina	59
A Brief History of Retail	60
Retail Eras	60
Aristide Boucicaut	61
The Shift	62

---

The Data Era	63
Stitch Fix	63
Keaton Row	65
Zara	66
Karolina (Continued)	67
<b>Chapter 5: Transforming Customer Relationships with Data</b>	<b>69</b>
Buying a House	69
Brief History of Customer Service	70
Customer Service Over Time	70
Boeing	72
Financial Services	74
The Data Era	75
An Automobile Manufacturer	76
Zendesk	76
Buying a House (Continued)	77
<b>Chapter 6: Intelligent Machines</b>	<b>79</b>
Denmark	79
Intelligent Machines	80
Machine Data	81
The Data Era	82
General Electric	82
Drones	84
Tesla	86
Networks of Data	87
Denmark (Continued)	88
<b>Chapter 7: Government and Society</b>	<b>89</b>
Egypt, 2011	89
Social Media	90
Intelligence	90
Snowden Effect	91
Privacy Risk Versus Reward	91
Observation or Surveillance	93
Development Targets	93
Open Data	95
Hackathons	95
Open Access	95
Ensuring Personal Protection	96
Private Clouds	97
Sanitizing Data	97
Evidence-Based Policy	97
Public-Private Partnerships	98

Impact Bonds	101
Social Impact Bond	102
Development Impact Bonds	103
The Role of Big Data	104
Egypt, 2011 (Continued)	105
<b>Chapter 8: Corporate Sustainability</b>	<b>107</b>
City of London	107
Global Megaforces	109
Population	109
Carbon Footprint	110
Water Scarcity	110
Environmental Risk	111
BP and Exxon Mobile	111
Early Warning Systems	112
Social Media	113
Risk and Resilience	114
Measuring Sustainability	115
Long-Term Decision Making	116
Stranded Assets	117
City of London (Continued)	118
<b>Chapter 9: Weather and Energy</b>	<b>119</b>
India, 2012	119
The Weather	120
Forecasting the Weather	120
When are Weather Forecasts Wrong?	121
Chaos	122
Ensemble Forecasts	122
Communication	123
Renewable Energy	124
Solar, Hydro, and Wind Power	124
Volatile or Intermittent Supply	125
Energy Consumption	126
Smart Meters	127
Intelligent Demand-Side Management	128
India, 2012 (Continued)	129

## **PART II: LEARNING FROM PATTERNS IN BIG DATA 131**

<b>Chapter 10: Pattern Recognition</b>	<b>133</b>
Elements of Success Rhyme	133
Pattern Recognition: A Gift or Trap?	134

---

What Fish Teach Us About Pattern Recognition	135
Bayes' Theorem	135
Tsukiji Market	135
Pattern Recognition	137
Rochester Institute of Technology	137
A Method for Recognizing Patterns	137
Elements of Success Rhyme (Continued)	140
<b>Chapter 11: Why Patterns in Big Data Have Emerged</b>	<b>141</b>
Meatpacking District	141
Business Models in the Data Era	142
Data as a Competitive Advantage	143
Data Improves Existing Products or Services	145
Data as the Product	145
Dun & Bradstreet	146
CoStar	148
IHS	149
Meatpacking District (Continued)	151
<b>Chapter 12: Patterns in Big Data</b>	<b>153</b>
The Data Factor	154
Summary of Big Data Patterns	155
Redefining a Skilled Worker	155
Creating and Utilizing New Sources of Data	156
Building New Data Applications	157
Transforming and Creating New	
Business Processes	157
Data Collection for Competitive Advantage	158
Exposing Opinion-Based Biases	159
Real-Time Monitoring and Decision Making	159
Social Networks Leveraging and Creating Data	160
Deconstructing the Value Chain	161
New Product Offerings	161
Building for Customers Instead of Markets	162
Tradeoff Between Privacy and Insight	163
Changing the Definition of a Product	163
Inverting the Search Paradigm for Data Discovery	164
Data Security	165
New Partnerships Founded on Data	165
Shortening the Innovation Lifecycle	166
Defining New Channels to Market	166
New Economic Models	167
Forecasting and Predicting Future Events	168

Changing Incentives	168
New Partnerships (Public/Private)	169
Real-Time Monitoring and Decision Making (Early Warning Systems)	169
A Framework for Big Data Patterns	170

## **PART III: LEADING THE REVOLUTION 171**

<b>Chapter 13: The Data Opportunity</b>	<b>173</b>
What Oil Teaches Us About Data	173
Bain Study	175
Seizing the Opportunity	176
<b>Chapter 14: Porsche</b>	<b>177</b>
Rome	177
Ferdinand Porsche	178
The Birth of Porsche	178
The Porsche Sports Car	179
Porsche Today	180
Rome (Continued)	180
<b>Chapter 15: Puma</b>	<b>181</b>
Herzogenaurach	181
Advertising Wars	182
Jochen Zeitz	182
Environmental Profit and Loss	183
Herzogenaurach (Continued)	184
<b>Chapter 16: A Methodology for Applying Big Data Patterns</b>	<b>185</b>
Introduction	185
The Method	186
Step 1: Understand Data Assets	187
The Patterns	188
Step 2: Explore Data	191
Challenges	192
Questions	192
Hypotheses	193
Data	193
Models	193
Statistical Significance	194
Step 3: Design the Future	194
The Patterns	195
Step 4: Design a Data-Driven Business Model	197
The Patterns	197



---

Step 5: Transform Business Processes for the Data Era	199
The Patterns	199
Step 6: Design for Governance and Security	201
The Patterns	201
Step 7: Share Metrics and Incentives	202
<b>Chapter 17: Big Data Architecture</b>	<b>205</b>
Introduction	205
Architect for the Future	206
Lessons from Stuttgart	207
Big Data Reference Architectures	207
Leveraging Investments in Architecture	208
Big Data Reference Architectures	211
Business View	212
Logical View	213
<b>Chapter 18: Business View Reference Architecture</b>	<b>215</b>
Introduction	215
Men's Trunk: A Retailer in the Data Era	216
The Business View Reference Architecture	217
Answer Fabric	218
Data Virtualization	219
Data Engines	220
Management	221
Data Governance	221
User Interface, Applications, and Business Processes	222
Summary	222
<b>Chapter 19: Logical View Reference Architecture</b>	<b>223</b>
Introduction	223
Men's Trunk: A Retailer in the Data Era (Continued)	224
The Logical View Reference Architecture	226
Data Ingest	227
Analytics	227
Discovery	228
Landing	228
Operational Warehouse	229
Information Insight	230
Operational Data	231
Governance	231
Men's Trunk: A Retailer in the Data Era (Continued)	232

<b>Chapter 20: The Architecture of the Future</b>	<b>233</b>
Men's Trunk: A Retailer in the Data Era (Continued)	233
Men's Trunk: Applying the Methodology	235
Step 1: Understand Data Assets	235
Step 2: Explore the Data	236
Step 3: Design the Future	237
Step 4: Design a Data-Driven Business Model	237
Step 5: Transform Business Processes for the Data Era	237
Step 6: Design for Governance and Security	237
Step 7: Share Metrics and Incentives	238
Men's Trunk: The Business View Reference Architecture	239
Answer Fabric	240
Data Virtualization	241
Data Engines	241
Management	242
Data Governance	242
User Interface, Applications, and Business Processes	243
Men's Trunk: The Logical View Reference Architecture Approach	244
Men's Trunk: A Retailer in the Data Era (Continued)	248
<b>Epilogue</b>	<b>249</b>
The Time is Now	249
Taking Action	250
Fear not Usual Competitors	251
The Future	252
 <b>INDEX</b>	 <b>255</b>

# Prologue

## BERKELEY, 1930s

George Dantzig sat in his dorm room, contemplating the next 24 hours and what it would mean for his future. He came to the University of California, Berkeley, with many aspirations, but as often happens, life got in the way, and his best laid plans turned into dreams for another day. As he gazed over the building immediately in the foreground, he could see Sather Tower on Berkeley's campus, known for resembling Campanile di San Marco in Venice. George reassured himself that one of his major goals was still in his grasp; he could still earn a position on the faculty, providing an opportunity to teach the next group of eager students.

It was 3 p.m., and George had until 8 a.m. the next morning to prepare for what would become his defining exam at Berkeley. A passing grade virtually guaranteed his spot on the faculty. Anything less than his best, and his future would be once again uncertain. This was the kind of motivator that got him to reopen the books and apply himself throughout the night. The last time George looked up from his book, he saw 3 a.m. on the clock and decided he should get some rest.

As the sunrise slowly emanated around his room, George opened one eye and then the other, immediately wondering why he had not heard his alarm yet. He figured it must be an exceptionally clear day for that type of light to be coming through his eyelids before the 7:15 a.m. alarm that he set. Suddenly, George felt like something was not right, sat straight up in bed, grabbed his glasses, and looked at the clock: 8:30 a.m. The exam began 30 minutes ago! George quickly pulled on his pants and dashed for the door.

George sprinted into the exam hall, where the professor greeted him with a surprised look. Obviously, the professor concluded that George must be in the hospital or perhaps even dead to have missed the start of the exam. George, in a rushed voice, explained the situation as his professor handed him the exam. He also noted, "George, there are three additional problems that I have written on the board, once you complete the questions on the exam paper."

George, without any minutes to waste, sat in the front row and quickly started working through the questions. The exam was set for three hours, so when George arrived at 8:50 a.m., many of the students were nearly halfway through with the questions. Two hours later, as the clock approached 11 a.m., George finished the last question on the paper exam and shifted his attention to the three questions on the board. George was the only student left in the hall and clearly he would not have a chance to finish. He sheepishly walked up to his professor, re-explained the situation, and apologized that he did not have time to answer the questions on the board. In an unexpected act, his professor offered to let George have until midnight to try to complete the additional questions, and George excitedly ran back to his dorm room.

It was now 3 p.m., over 24 hours since he reflected on his future in his dorm room. George made progress on one of the questions but decided to give up on the others. He spent the next eight hours grinding on the first question, feeling confident that he had conquered the problem, and set out across campus to turn in his single answer. His disappointment was obvious in his posture — while he felt a great sense of accomplishment on one question, he knew that providing only one answer out of three would not be sufficient. George slid the paper under his professor's door, grabbed a small bite at the campus cafeteria, and collapsed into his bed at 1 a.m.

George was awakened by the shrill sound of his phone at 7 a.m., and he heard his professor on the line, "George, I can't believe it. You actually solved one of the equations on the board! This is truly an historic day." George, confused by what he was hearing, asked his professor to explain his amazement. His professor replied, "George, when I handed out the test, I told the class that I wrote three unsolvable equations on the board. I expected anyone with extra time to play around with them, but they weren't actually part of the test. You accomplished something that the rest of us KNEW was impossible!"

It's amazing what we can accomplish when we are not encumbered by what we believe is possible. It turns out that George had solved an algorithm around linear programming, which eventually became the simplex algorithm, the heart of Microsoft Excel's SOLVER function. If George had arrived on time to the exam hall and heard his professor tell the class to try to solve the unsolvable problems, he probably would have never accomplished this feat. *He was not limited by what the world felt was impossible.*

## PATTERN RECOGNITION

In most great detective stories, the author typically uses a signature act or prop. The robber or murderer, either intentionally or unintentionally, leaves a signature mark at the scene of a crime. The detectives in pursuit do not notice it at first but eventually come to see a pattern. Often, it's about

observing a non-obvious relationship between two data points, which leads to the recognition of a pattern. The conclusion: Insight can be found anywhere and hides in patterns. But are patterns valuable only to Sherlock Holmes and his contemporaries?

Patterns exist everywhere we look. However, it is the actual recognition of the pattern that can truly impact the world in which we live. When studying highly successful people, whether they're athletes or successful business professionals, Malcolm Gladwell argues in his best-selling book *Outliers* (Little, Brown and Company) that by studying the patterns of these individuals, we begin to understand that the environment from which they come from directly correlates to the amount of success they will have in life. By looking at hidden patterns of these individuals' upbringings, the month they were born, or the culture in which they were raised, we can predict whether they will reach their full potential or perhaps fall short.

The same can be said for data. By capturing and analyzing data, we can find patterns that will ultimately impact the future of industries and businesses. Often, observing a non-obvious relationship between two data points leads to the recognition of a pattern. The conclusion? Insight can be found anywhere and hides in patterns.

## NELSON PELTZ

Nelson Peltz's coffee mug on his desk reads Sales Up . . . Expenses Down on one side, with Cash Is King on the other. These mantras helped him to build an \$8.5-billion partnership focused on activist investing, one of the largest such partnerships in existence today. However, this was not handed to him; Peltz built this partnership through his ability to see, understand, and apply patterns across a variety of businesses.

Born in 1942 in Brooklyn, New York, Peltz worked his way through early school years and eventually decided to attend the Wharton School at the University of Pennsylvania. He dropped out a couple years later, in 1963, and set off to be a ski instructor. When that did not work out, he returned home to drive a delivery truck for his grandfather's company, A. Peltz & Sons. Eventually, he was given the reins to the company (Flagstaff) and grew it into a publicly held company. While Nelson did not finish a formal education, his on-the-job learning taught him all he would need to know later in his life.

In the 1980s, Nelson reunited with a former business partner from Flagstaff, Peter May, and they went on a hunting expedition: looking for companies to acquire, grow, and eventually sell. His first marked success was the sale of Triangle Industries to Pechiney in 1988. Nelson began to notice the value of spotting patterns in the operations of the companies that he acquired or considered acquiring.

Eventually, Trian Fund Management was founded, and still exists today. He became a force advocating for change and was an activist investor before the term was en vogue. But how did he do it?

Peltz believes that most activist investors and private equity companies focus on financial engineering. While that can make a difference in many cases, it gets you only so far. Very few individuals possess the insight to improve business operations, which is where you can drive substantial returns. And, in Peltz's view, improving business operations requires identifying and understanding patterns, and then acting on them.

To drive meaningful insight out of patterns, Peltz consciously (or unconsciously) focused much of his investment in two sectors: consumer packaged goods and food. With a focused competence, he can better compare and contrast patterns of performance. He believes in constantly assessing companies in his area of competence and looking at their strengths and weaknesses against the best performers in the industry. A sampling of the patterns that he looks for include:

- **Ratios:** Percent of sales spent on marketing.
- **Overhead:** Growth in overhead versus growth in sales.
- **Rebates and allowances:** Deals and allowances paid to retailers.
- **Brand:** He favors companies that have strong brands, but also have ratios that are out of line with the best performers.

Peltz believes that it's a more efficient use of capital to revitalize previously great brands than to try to build a new one. This is why he began to take an active interest in Heinz and Wendy's.

## Heinz

When Peltz began analyzing H.J. Heinz Company in 2005, it was an immediate fit to his pattern-focused investing style. He saw a company with brand value and strong free cash flow, yet the total shareholder returns trailed the S&P 500, the large-cap food index, and the mid-cap food index. But why?

He observed that Heinz's Selling, General, and Administrative (SG&A) expenses, as a percent of revenue, was dramatically higher than the best comparable performers. The advertising investment as a percent of revenue was also out of line. Next, he noticed that the rebates and allowances being paid to retailers were much higher than the amounts paid by other organizations in its peer group. In his mind, that money could be put into marketing and product innovation, instead of lining the pockets of the retail channel. He noticed that plant efficiency metrics also trailed the best performers.

Lastly, he highlighted the fact that all of Heinz's businesses were operating at margins in excess of the company average, indicating that the overhead at headquarters was crippling the business. All the patterns that he had learned to watch for over time were present at Heinz.

## Wendy's

Wendy's, like Heinz, demonstrated a pattern of underperformance versus its peer group. In the case of Wendy's, through a ratio analysis, Peltz could see that margins were unacceptably low (10 percentage points below its peer group) and driven by excessive overhead and operating costs. Next, Wendy's lost its focus on brand strength, as it had diversified into other food categories like Tim Hortons Cafe and Bake Shop in Canada. Different company, same patterns.

While it seems simple to observe after the fact, these were new insights, previously unnoticed at the time.

Peltz has gone on to advocate for similar change in the likes of Cadbury-Schweppes, Kraft Foods, Snapple Beverage Corp., PepsiCo, and many others. All these companies fell within his core competence, all assessed against the same set of patterns, and each one was driving value for Peltz, his investors, and the investors in those companies. In the last few years, the assets under management at Triun increased from \$3.7 billion in 2012, to \$6.3 billion in 2013, to \$8.5 billion in the most recent report.

The power of being able to identify, understand, and execute upon patterns of success is critical in the pursuit of distinguishing oneself or an enterprise.

## COMMITTING TO ONE PERCENT

Dave Brailsford basked in the glow of the Olympics. The Great Britain cycling team just completed their participation at the 2012 Olympics in London, England, winning 70 percent of the medals in men's cycling. Reporters probed with aggressive questioning, wanting to understand the silver bullet that led to this success. The irony: There was no silver bullet. In fact, it was the opposite of a silver bullet.

When the Great Britain cycling team, Team Sky, hired coach Dave Brailsford in 2010, the country had never won a Tour De France. In fact, the history of the sport in the country was filled with errors, mishaps, and minimal success. Historically, the team chased fads of success: new equipment, new uniforms, new techniques. But nothing changed the trajectory. Then, Dave Brailsford arrived.

Dave Brailsford fanatically talks about the aggregation of marginal gains. This concept means that by marginally improving each and every aspect of a process, the aggregation of those small gains will lead to large improvements. Brailsford's goal was simple: one percent. He sought a one-percent improvement in every aspect of the cycling team.

Setting out to improve all aspects of a cycling team, the obvious places to start are in areas like nutrition, bike performance, and physical conditioning. After all, improving every meal by one percent promised a path to continued improvement. However, for Brailsford, those enhancements merely scratched the surface. He set out to improve *every* aspect by one percent. Not only sports massage, but the gels used for sports massage. Not only the bikes, but the grips on the bikes and, more specifically, the tackiness of the grips. He studied not only the physical conditioning, but also the sleep habits and, more specifically, the pillows used. He focused on every aspect: one-percent improvement. It's that simple.

In 2012, a short two years after Brailsford joined the team, Great Britain won its first Tour De France. Shortly thereafter, the triumph at the Olympics in London occurred. The aggregation of the one-percent gains created superior outcomes.

## THE BIG DATA REVOLUTION

The big data revolution is about accomplishing feats with data that no one believes is possible. The leaders of the big data revolution will embody three characteristics:

- The ability to suspend disbelief of what is possible, and to create their own definition of possible
- An inherent knowledge of pattern recognition and the insight to apply patterns from one industry or dimension to another that may be seemingly unrelated
- Commitment to one-percent improvement in every aspect related to data

Combining these seemingly different characteristics and applying them creates possibilities previously undetected. By empowering ourselves with data and believing we can discover the undiscovered, we can launch businesses and industries to new levels. Throughout this book, we will look at different industries and how they use data to make their impossible possible.

This revolution is about finding your possible.



# Introduction

## STORYTELLING

Stories make for powerful communication, and this book is a compilation of the stories, patterns, and methods that we have seen over the last decade, since we've been focused on this emerging phenomenon: Big Data. The stories are based on true events but do not always include actual names, events, or circumstances.

These stories are meant to illustrate the challenges and possibilities present with the advent of big data, based on what we have witnessed. Our belief is that such stories provide the best way to learn about how other business leaders both responded to external change and in some cases caused disruptive change within their respective industries. We hope that they will provide a source of inspiration, courage, and know-how, so that you can embrace big data as a means of inciting a revolution within your organization.

## OBJECTIVE

The target audience of this book ranges from the entrepreneur, to management in established enterprises, to those who are merely curious about the implications of big data in their own lives. This book is intended to provide ammunition for breaking down the barriers that often exist between those who manage data and those who manage people. Data is the new intellectual property. It can be harnessed for advantage or ignored at peril.

For those who would prefer to remain working in silos, where data analysis and decision-making is divorced, this book will make for uncomfortable reading. Organizations that do not manage to utilize their data assets will eventually become extinct. The challenge of improving connectivity between data and behavior, and between machine and human, will require dedicated effort in terms of building human capacity and financial resources. In addition, patience is required while organizations make this transition. For some organizations, it may be necessary to initiate external activities or form partnerships in order to adequately assess the potential value of big data.

Unfortunately, many have a vested interest in resisting the data revolution due to their fears about the impact it will have on their own professions. It is likely that such resistance will be futile and that those who actively embrace the oncoming disruptive change will benefit most from the opportunities offered. Estimates by C.B. Frey and M.A. Osborne (“*The Future of Employment*” [Oxford University, 2013]) suggest that almost half of existing jobs will be at risk of automation due to the technological developments that are currently taking place because of big data and the application of machine-learning approaches. While many routine tasks are already being computerized and automated, recent scientific advances suggest that it will be possible to automate an increasing amount of non-routine cognitive tasks, such as accountancy, legal work, technical writing, and many other white-collar occupations.

We argue throughout the book that in order to develop strategies for managing organizations in a knowledge-based society, it will be necessary to grab hold of data opportunities before agile, data-savvy competitors pass you by. The transition needed is complex and involves changes to technology, processes, and human behavior. There is no doubt that surviving such disruptive change will be a substantial challenge for many. While there is no magic formula to fit all types and sizes of organizations, we offer a strategy in the form of a roadmap that can be easily tailored to suit each individual organization.

Our approach to creating a roadmap for success in the Data era is to first explain how other organizations have changed and outsmarted their competitors. The first part of the book presents nine stories about innovation using data. The second part of the book gathers together the key patterns that we have observed through the stories presented. By distilling these patterns, we offer a broad understanding about how data is being leveraged across a range of different industries. Finally, in the third part of the book, we aim to inspire business managers to lead the revolution by offering a methodology for operationalizing big-data approaches that can be adapted for different industries.

## OUTLINE

### PART I “THE REVOLUTION STARTS NOW: 9 INDUSTRIES TRANSFORMING WITH DATA”

Part I shares the main stories that will be used to illustrate patterns and concepts throughout the book. While we had hundreds of stories to choose from based on our experiences, we chose the nine that we think are the most compelling and engaging, and that bring out the best illustration of the

patterns that we have seen. The stories can be read in any order, so we suggest you pick the one that sounds the most interesting, and then work your way through the others:

- **Chapter 1: Transforming Farms with Data.** Rob starts with a quick look at the history of farming and the evolution of technology in agriculture. The chapter delves into limitations through the years and how they have been overcome. The majority of the chapter explores how data will continue to transform this seemingly analog industry. As you will see, many players in this industry have not yet awakened to the impact of data, and they are quickly being passed by, perhaps without even noticing.
- **Chapter 2: Why Doctors Will Have Math Degrees.** Historical approaches to medicine, treatments, and wellness are not relevant in the data era. The chapter highlights how many decisions in the medical field today are based on opinions instead of facts, and how this leads to suboptimal outcomes. We also showcase a new set of leaders in medicine who are disrupting traditional industry practices through the use of data. The key implication is that the role and skills of doctors will change in the Data era.
- **Chapter 3: Revolutionizing Insurance: Why Actuaries Will Become Data Scientists.** The insurance industry is undergoing a fundamental shift based on better collection, access, and usage of data. Underwriting and actuarial services, which are largely about forecasting what might happen, will take a backseat in a world where you can monitor what is actually happening and price accordingly. New business models are emerging, which is disrupting the traditional skills and tools needed to win in insurance.
- **Chapter 4: Personalizing Retail and Fashion.** Data can turn traditional segment-based retailing into a more personal approach: thousands of individual customers, instead of thousands of customers. There is a timeless quote in retail stating, “I know that only half of my marketing is effective. The problem is that I don’t know which half is working.” Transforming retail, however, is more than just using data to better target clients. It’s about using data to transform the role of a retailer and truly serve a customer of one.
- **Chapter 5: Transforming Customer Relationships with Data.** Data will increase the intimacy between the firm and the customer. By improving the collection, relevance, and utilization of data about customers, firms will be able to maximize customer satisfaction by processing data about individuals in real-time. Rather than responding to issues and problems, data about the locations and preferences of individuals will allow organizations to offer services and solutions to improve their personal experiences and identify challenges ahead of time.

- **Chapter 6: Intelligent Machines.** Rob starts with a story about visiting Denmark and his exploration of the wind turbine business. The chapter goes on to discuss how previously unconnected machines are coming to life through connectivity and data. Bringing life to machines may seem futuristic, but there are already many developments that are facilitating the production of machine-readable data that is helping to increase intelligence. The Internet of Things describes the network of such machines and their ability to share information. The Industrial Internet heralded by General Electric, intelligent wind turbines, the potential of drones, and Tesla's Vehicle Management System are all examples of how networks of data will revolutionize our world.
- **Chapter 7: Government and Society.** Closing the loop between people and government using data has considerable potential. Whereas elections, referendums, and opinion surveys cost substantial amounts of money, social media offers a means of monitoring public opinion, assessing perceptions, and testing and fine-tuning public policy. At the same time, privacy risk is now a major concern in many countries and is delaying data open-access initiatives. Finding a reliable way to address these risks through anonymization techniques without degrading the quality of the data will be a challenge. This chapter also explores the potential rewards of using big data for public-private partnerships for delivering socio-economic benefits.
- **Chapter 8: Corporate Sustainability.** Connecting people through the Internet and social media has increased awareness about the global supply chain behind many of the services and products that we consume on a daily basis. Faced with the collective responsibility for ensuring sustainable practices, many firms are now seeking to become leaders within their industry and are also reaping the benefits of moving first. Having the confidence to design and implement a corporate sustainability strategy requires the capability to assess the risks associated with future scenarios. Agitation for change is coming from those that have the mandate to make long-term decisions.
- **Chapter 9: Weather and Energy.** Forecasting the weather has been and will continue to be a challenge, despite the many scientific advances that have been made in terms of data, models, and techniques. Nevertheless, weather forecasting serves to illustrate how human behavior relates to the task of generating and responding to future scenarios. The close relationship between weather and energy shows how big data will be used to operate power systems when substantial amounts of variable renewable energy are integrated. Although introducing many changes, the combination of better data and technology innovation will help to balance supply and demand and keep the lights on.

## PART II “LEARNING FROM PATTERNS IN BIG DATA”

The second part of the book distills the nine stories in Part I into a set of discrete patterns. We explore the concept of pattern recognition, how it can be applied in a multitude of settings, and the implications for the Data era. We close this part of the book with a detailed discussion of the 54 patterns in big data that we have observed:

- **Chapter 10: Pattern Recognition.** Identifying patterns is usually the first step to constructing systems for forecasting, classifying, or simply segmenting customers. The ability to identify patterns in big data, and to also assign significance levels to these patterns, will be extremely important in the Data era. Collecting relevant data, cleansing this data, extracting appropriate features, classifying the data, and evaluating confidence is all part of the process. From understanding Bayes theorem to distinguishing species in a Tokyo fish market, pattern recognition algorithms vary from human intuition to sophisticated machine-learning algorithms.
- **Chapter 11: Why Patterns in Big Data Have Emerged.** Emerging patterns in applications of big data in different organizations can help illustrate the potential of this revolution. There are three prominent approaches to building a business model in the Data era. In some cases, data provides a competitive edge by moving before the crowd. In others, it improves the existing products or services. Finally, data can become the product itself by recognizing the need for firms to obtain access to particular datasets.
- **Chapter 12: Patterns in Big Data.** Identifying patterns in big data can help anyone prepare their organization for the Data era. A series of 18 data factors are described, based on the stories from Part I of the book. These factors are then further decomposed into 54 big-data patterns, with the aim of representing the best practices of a range of leading-edge firms in the Data era.

## PART III “LEADING THE REVOLUTION”

The third part of the book focuses on how to create a big-data revolution in your own organization. You need to develop an appreciation for the lessons learned from the individual stories in the first part of the book, coupled with a bias for action. We recommend thoughtful consideration about how the patterns that have been extracted in Part II of the book apply (or don't) to your organization. The substantial challenge facing each individual business leader is to determine the steps necessary for operationalizing the required changes in as short a time as possible:

- **Chapter 13: The Data Opportunity.** Focusing on opportunity is important to encourage behavior change. Similarities between the processes of finding, refining, and adding value for two commodities, oil and data, help to illustrate how value will be created in the future. Empirical support for the benefits of a data focused strategy first come from a Bain report published in 2013. Early adopters of big data are twice as likely to be in the top quartile of financial performance within their industries.
- **Chapter 14: Porsche.** Aiming for perfection is key to success. Rob begins with a story set in Italy, and expands into a story about Ferdinand Porsche, the founder of Porsche. The story serves to demonstrate why innovation, adaptability, and perseverance are responsible for the success of this impressive sports-car business. Porsche's fit-for-purpose approach ensures that their vehicles are designed to deliver to many different client needs while optimizing performance, quality, and value. A similar approach, fit-for-purpose, will be taken by the leaders of the Data era.
- **Chapter 15: Puma.** Striving to be the best is one of the hallmarks of a successful organization. This story of the aggressive battle between Puma and Adidas is rooted in sibling rivalry and went well beyond business competition. Jochen Zeitz took over the reigns of Puma in 1993 and delivered a 4,000-percent increase in Puma's share price over the next 13 years. Recognized as a great business leader, Zeitz has used big-data analytics to create an environmental profit-and-loss accounting system and is now advocating its use in other firms.
- **Chapter 16: A Methodology for Applying Big-Data Patterns.** This methodology, divided into seven specific steps, provides a recommended approach for applying big-data patterns in any organization. While the methodology is intended to be applied sequentially, some organizations may have already completed (or at least started) some of the steps. The methodology should be used as a roadmap, as opposed to a destination: Use the parts that you need.
- **Chapter 17: Big-Data Architecture.** As an organization works through the methodology in Chapter 16, the thinking will often turn to execution. Chapter 17 takes the first steps towards execution, sharing the landscape of a big data reference architecture. The focus is on Business View and Logical View reference architectures.
- **Chapter 18: Business View Reference Architecture.** This chapter dives into the components of a Business View reference architecture. We introduce you to a fictional retailer, Men's Trunk, which illustrates a journey into the Data era. The Business View reference architecture includes the Answer Fabric, Data Virtualization, Data Engines, Management, Data Governance, and User Interface/Applications components.