Open Data Now

The Secret to Hot Startups, Smart Investing, Savvy Marketing, and Fast Innovation

Joel Gurin
Contents

Preface: My Journey Through the Datashere ix
Acknowledgments xiii
Introduction: The World’s Greatest Free Resource 1

Part 1 The Power of Open Data
1 • An Opportunity as Big as the Web 7
2 • Hot Startups: Turning Government Data into Dollars 23
3 • Consumer Websites: Choice Engines for Smart Disclosure 43
4 • New Companies to Manage the Data Deluge 67
5 • Data-Driven Investing: New Tools for Business Analysis 81
6 • Green Investing: Betting on Sustainability Data 93
CONTENTS

7 • Savvy Marketing: How Reputational Data Defines Your Brand 111
8 • The Marketing Science of Sentiment Analysis 129
9 • Tapping the Crowd for Fast Innovation 141
10 • The Open Research Lab: Innovating Through Open Collaboration 155

Part 2 The Business Environment: New Trends in Open Data
11 • Privacy, Security, and the Value of Personal Data 177
12 • Doing Business in a See-Through Society 197
13 • Government and Data: Setting the Rules for an Open World 213
14 • The Open Data Future 235

Selected Bibliography 245
Appendix A: Using OpenDataNow.com 251
Appendix B: Defining Data Categories 253
Glossary 257
Chapter Notes 275
Index 313
I’ve been fortunate to have a varied and satisfying career. I began as a science journalist; became a book author and magazine editor; was the editorial director and then executive vice president of Consumer Reports for many years; worked in Washington as a bureau chief at a federal agency; and held leadership positions in publishing and nonprofit organizations in between. But nothing has been more fascinating and inspiring than the work I’ve had a chance to do on Open Data for the past three years.

I’ve been immersed in the world of Open Data since the fall of 2010. I had joined the federal government almost a year before as chief of the Consumer and Governmental Affairs Bureau at the Federal Communications Commission. I was recruited to join the FCC
because I had a strong consumer background and the new chairman, Julius Genachowski, wanted to make consumer issues a priority. As soon as I joined the agency, my bureau and I started to tackle a thorny consumer problem: How could we help people choose from a dizzying array of cell-phone plans to find their best personal choice?

To try to figure that out, I started talking to colleagues who were working on similar consumer problems in financial services, healthcare, and other areas. Eventually I met Danny Goroff, then at the Office of Science and Technology Policy, who suggested that someone should start a White House task force to look at this issue. The result: I became chair of the White House Task Force on Smart Disclosure, cochaired by Sophie Raseman of the Department of the Treasury, which we launched to study how consumers can use government Open Data about products and services to make more informed choices in the marketplace.

Through the task force I met Beth Noveck, who had led the Open Government Initiative for the Obama administration. After working with her on several projects, I joined her team as senior advisor at the Governance Lab at New York University. The GovLab, as we call it, develops new approaches to government through technology, data, and collaboration, and Open Data is a central part of our work.

I’ve gotten to know dozens of entrepreneurs, business leaders, researchers, academics, journalists, and others who are developing new ways to understand and use Open Data. I’ve met a new breed of techno-evangelists whose rallying cry is “data liberation!” And I’ve discovered a growing, vibrant new world of Open Data resources, a kind of datasphere that provides the material for transformational changes in business, government, and society.

I’ve been lucky in my career to work in the private sector, in consumer advocacy, and in government, and this book reflects all three perspectives. I hope it will be useful to business readers, the primary audience for this book, who want to grow their companies, market
more effectively to their customers, find new investment opportunities, and build new businesses. But I also hope it’s helpful to the government agencies that work with those businesses, to the advocates who keep both business and government honest, and to consumers who want to use Open Data to improve their own lives.
Open Data is “going to help launch more startups. It’s going to help launch more businesses. . . . It’s going to help more entrepreneurs come up with products and services that we haven’t even imagined yet”—President Barack Obama, May 9, 2013.

Imagine that you suddenly have access to a massive new business resource. It can be used to build new companies that deliver better healthcare, offer in-depth investment tools, provide energy more efficiently, improve transportation, or offer a wide range of consumer services through mobile apps and the web. It can help established companies learn what their customers really think about them, spot market trends early on, and choose the best business partners while avoiding the riskiest ones. It can speed up scientific research by an order of magnitude. And while it takes money and effort to develop these opportunities, the essential resource itself is free.
That resource is Open Data. The concept of Open Data is simple enough: it's public data, from government or other sources, that's available for anyone to access for personal or business use. Anyone who’s ever booked a flight online, used a smartphone’s GPS, or watched the Weather Channel has used Open Data. This free public data can also be used, with ever-improving methods of data analysis, to launch new business ventures, solve difficult strategic problems, provide new kinds of business intelligence, and more.

Today’s Open Data revolution is rapidly leading us into new territory. With more powerful computers, cheaper memory storage, and the exponential growth of digital information, huge databases are now becoming public for the first time. They are taking Open Data to an entirely new level and changing our world in the process.

Open Data is becoming a secret to success for smart business leaders around the world. Investors use it to analyze the risks and rewards of different companies in their search for the best opportunities. Company owners use it to understand subtle clues to their brand’s reputation and to develop data-driven marketing strategies. And entrepreneurs are using Open Data on weather, housing, transportation, and more to build businesses that provide new services and benefit the public. These startups are raising tens of millions of dollars and can build value many times greater than that.

In October 2013, McKinsey & Company released a new report on the value of Open Data worldwide. The report calculated that “an estimated $3 trillion in annual economic potential could be unlocked” through Open Data in sectors like education, energy, and health. This is the largest value put on Open Data so far, and may seem too inclusive; for example, it includes the value of using Open Data to increase the population’s earning power by improving education. Regardless, this latest McKinsey report demonstrates the growing business interest in Open Data and excitement about its potential.

At the same time, Open Data is having a profound impact outside the private sector in ways that affect the business environment. Government
leaders are releasing more Open Data both to enhance trust in government and to foster business innovation. Medical researchers are using it to find treatments and cures for diseases more rapidly. Journalists use Open Data to expose problems ranging from money laundering to Medicare fraud—and increasingly to bring corporate operations to light.

Over the past three years I’ve met with the leaders of the emerging Open Data movement, first as head of the consumer bureau at the Federal Communications Commission, then as chair of a White House task force, and now as part of a team at New York University. I’ve met with inspiring entrepreneurs who are using Open Data to build new enterprises and drive progress in areas like healthcare, finance, education, and energy. I’ve gotten to know high-level U.S. and U.K. government officials, policy makers, academics, and NGOs who are working to release all kinds of new data that businesses and others can use. This book reflects their vision, insights, knowledge, and advice.

The world of Open Data is moving fast, and no book on this topic can be completely current. I’ve launched a website, OpenDataNow.com, to follow the latest developments, debates, and opportunities. The site hosts a blog, apps, and news links to serve as a hub for the growing Open Data community. I encourage readers to check it out, comment, and contribute. I’ve posted articles, interview transcripts, and podcasts that focus on several of the experts quoted in this book; a list of these is in Appendix A on pages 251-252.

This book contains several resources to help readers explore the possibilities of Open Data. Most chapters conclude with pragmatic advice on putting Open Data to use. The book also includes an extensive Glossary that should help anyone new to this field. (I could have used it myself when I was doing this research.)

Whether you are an entrepreneur, an executive, an investor, a small business owner, or in the public or nonprofit sector, Open Data will provide new opportunities that could transform the way you work. *Open Data Now* is your guide to understanding this new phenomenon, managing its challenges, and harnessing its power.
PART ONE

The Power of Open Data
In November 2012 I was sitting in a packed conference room at the brand-new Open Data Institute (ODI) in London, a public-private partnership launched with a 10-million-pound grant from the British government. The ODI has all the look and feel of a well-funded tech startup. The Institute is situated in London’s equivalent of Silicon Valley, known as Silicon Roundabout for the Old Street roundabout where the nearest tube station is. It’s an area much like the tech hives in San Francisco’s South of Market neighborhood or Silicon Alley in lower Manhattan—underdeveloped neighborhoods with ample space for new companies to stake out their territory, innovate like mad, and drive up the cost of real estate. While this international gathering happened to be in London, it could have been in either of those technology centers.

I was at the ODI with two dozen colleagues from business, government, and nonprofit organizations who had come together to talk about Open Data—the new movement to make large amounts of data available for public use. On this November Monday, we were in
the second of two back-to-back meetings at the ODI, put together with funding from the MacArthur Foundation. We were meeting as representatives of the White House; 10 Downing Street; the U.K. government’s Cabinet Office and its Department of Business Development; the World Bank; a major British retailer; two high-tech consulting firms; a leading tech publisher; university departments of law, computer science, artificial intelligence, physics, and cognitive neuroscience; two foundations; and nonprofits working on corporate transparency, civic engagement, and green business practices.

We were greeted by the Institute’s CEO, Gavin Starks, an entrepreneur who has done Internet development for two decades in places as diverse as Virgin, Google, the British government, and UNICEF. He was making the provocative case that Open Data would have the same impact as the invention of the World Wide Web.

Open Data today, he said, “is very much like the web was for me in 1994, when I was still trying to convince people that e-mail was a good thing or that they might want to launch a website. Everyone was excited about its potential, but no one knew quite what shape it would take. Over the last 20 years we’ve seen a lot of innovation, creativity, and disruption. Today, we don’t know exactly where Open Data will lead, but we do know that it will be transformative. Some ways of doing business will start, some will evolve, and learning how to navigate that will be a challenge for all of us. But the potential that we saw in the early days of the web is what I see now with Open Data.”

This isn’t the first time a new tech development has been compared to the dawn of the web. But the leaders of the Open Data movement have the credentials to make that claim. The president and cofounder of the Open Data Institute is Sir Tim Berners-Lee, whose bio says, simply and accurately, that he “invented the World Wide Web” while at CERN, the European Particle Physics Laboratory. The ODI’s chairman, Sir Nigel Shadbolt, is a pioneer in web science and artificial intelligence and was instrumental in creating the British government’s Open Data policies. They are typical of the visionaries
in the United States, the United Kingdom, and other countries who are developing Open Data now.

Open Data can best be described as *accessible public data that people, companies, and organizations can use to launch new ventures, analyze patterns and trends, make data-driven decisions, and solve complex problems*. It's very different from Big Data—more on that in a minute—although the two overlap. Open Data is data with a mission: it’s designed to provide free, open, transparent data that can transform the way we do business, run government, and manage all kinds of transactions. Like our gathering at the ODI, the people behind Open Data are a diverse group, including leaders from the corporate world, technology, government, academia, nonprofits, and fields such as health, education, and environmental science.

The Open Data movement began with democratic goals, fueled by the idea that governments should make the data they collect available to the taxpayers who’ve paid to collect it. But in addition to its social benefits, Open Data has created tremendous new business opportunities, which are the focus of this book. It’s worth remembering that the Internet itself began as a government-funded initiative, the ARPA-NET, created by the Advanced Research Projects Agency that President Eisenhower launched as a response to Sputnik. That government research project became one of the major economic drivers of our time. In a similar way, government’s drive to release Open Data is creating a major economic resource and the infrastructure to manage it.

The Open Data policies developed by the U.S. and U.K. governments are driven by a push for economic growth and job creation. President Obama made this clear when he announced his administration’s new Open Data Policy in May 2013. This policy, which will make unprecedented amounts of federal data available in highly usable forms, has a business agenda first and foremost. Significantly, the president didn’t make his announcement at a Washington press conference or in the Rose Garden but on a visit to a technology center in Austin, Texas. There he promised that government Open Data
is going to help launch new businesses of all kinds in ways “that we haven’t even imagined yet.”

The Open Data Policy includes a detailed description of the criteria for government data to be released as Open Data, drawing on work done by the Open Knowledge Foundation in the United Kingdom, the Washington-based Sunlight Foundation, and others. This book goes further: the Open Data I’m writing about includes data from other sources as well as government.

I use “Open Data” to include data from any source that’s made available in an “open” form that anyone can access and that meets a few specific conditions. All Open Data must be licensed in a way that allows for its reuse. It should be in a form that can be easily read by computers, although here there are gradations of “openness.” And there’s general agreement that Open Data should be free of charge or cost just a minimal amount.

Open Data includes federal, state, and local data; scientific data released by researchers; data that companies release about their own operations; user reviews and tweets written by ordinary people; and any kind of data that can be found through Google or scraped from websites. By using these many kinds of Open Data:

- Entrepreneurs are building new businesses that generate many millions of dollars in revenue. Open Data released by the National Oceanic and Atmospheric Administration beginning in the 1970s and GPS data released more recently spawned new industries that do billions of dollars in business each year. New businesses using open health data may soon match that, and opportunities in energy, finance, education, and other fields are increasing as well.
- Governments are providing new, centralized data resources for business development. Data.gov, a website launched by the Obama administration, now makes hundreds of thousands of government datasets open and available for anyone to use for
free. The United Kingdom has launched its own version, Data.gov.uk, and other countries are using a platform distributed by the United States as “Data.gov in a box” to start their own data hubs.

- Companies are developing new marketing strategies, evaluating competitors and partners more accurately, and building their brands’ value. The new technique of *sentiment analysis* gathers information from Twitter, blogs, news feeds, and other public sources, uses text analysis to turn this information into Open Data, and turns the mass of public opinion into quantifiable business insights.

- Investors are finding companies with the greatest promise and avoiding those that pose high levels of risk. Through new data-driven websites, investors can quickly get in-depth information on large and small companies. Open Data is giving investors new insights into companies ranging from innovative startups to globally traded public corporations through websites that provide online tools and data visualizations.

- Companies are becoming more transparent about their operations, to their benefit. Between government-required disclosures and voluntary reporting, companies are making more Open Data available about their environmental, social, and governance practices. By releasing this data, a company can attract new investment, recruit more effectively, and improve its corporate image.

- Scientists and researchers are accelerating the pace of new discoveries. In the physical sciences and biomedicine, researchers are taking the bold step of both sharing their data early and openly so that online networks of both experts and amateurs can work with their data to achieve new breakthroughs. Even the secretive world of drug research is beginning to make more data public.

- Websites are helping consumers make better, more informed choices for all kinds of products and services. New businesses
are developing online and mobile “choice engines” that give consumers the data they need to make complex, important decisions. They help consumers access detailed, interactive Open Data to choose the options that are best for them, whether they’re choosing healthcare, a mortgage, a credit card, or a college education.

Open Data vs. Big Data: Related but Very Different

Open Data should not be confused with Big Data, one of the most talked-about developments in information science over the last few years. Big Data involves processing very large datasets to identify patterns and connections in the data. It’s made possible by the incredible amount of data that is generated, accumulated, and analyzed every day with the help of ever-increasing computer power and ever-cheaper data storage. It uses the “data exhaust” that all of us leave behind through our daily lives. Our mobile phones’ GPS systems report back on our location as we drive; credit card purchase records show what we buy and where; Google searches are tracked; smart meters in our homes record our energy usage. All are grist for the Big Data mill.

While Big Data and Open Data each have important commercial uses, they are very different in philosophy, goals, and practice. For example, large companies may use Big Data to analyze customer databases and target their marketing to individual customers, while they use Open Data for market intelligence and brand building. National governments may use Big Data to track citizens in the name of security, while they use Open Data to engage with their citizens and foster participatory democracy. It’s telling that the recent book Big Data, the best general presentation of the field, devotes only two-and-a-half pages to Open Data. The two are not the same.

With Big Data, the data sources are generally passive, and the data is often kept private. Big Data usually comes from sources that passively generate data without purpose, without direction, or without
even realizing that they’re creating it. And the companies and organizations that use Big Data usually keep the data private for business or security reasons. This includes the data that large retailers hold on customers’ buying habits, that hospitals hold about their patients, that banks hold about their credit card holders, or that government agencies collect about millions of cell-phone calls.

At this writing in the fall of 2013, I’ve found that every time I mention the word *data* it triggers a discussion about the National Security Agency and its PRISM program. We’re still trying to figure out exactly what data the NSA has collected, how much, and why. The NSA revelations have rekindled a national debate about data privacy, which is a good thing (more on that in Chapter 11). PRISM is a prime example of the disturbing side of Big Data: it’s a massive collection of data without the participation, or even the awareness, of the people whose data is being collected, and it’s been kept hidden from the public until recently. It’s also the antithesis of Open Data. In fact, even the idea of Open Data for national security is an oxymoron.

In contrast to most Big Data, Open Data is *public* and *purposeful*. It’s data that is consciously released in a way that anyone can access, analyze, and use as he or she sees fit. (I don’t count Edward Snowden’s revelations as Open Data; to be truly open, data should be released by someone who has the authority to do so, not by someone who has pilfered it.) Open Data is also often released with a specific purpose in mind—whether the goal is to spur research and development, fuel new businesses, improve public health and safety, or achieve any number of other objectives.

Having said all that, Big Data and Open Data do overlap, and when they do, the result can be powerful. Some government agencies have made very large amounts of data open with major economic benefits. National weather data and GPS data are the most often-cited examples. U.S. census data and data collected by the Securities and Exchange Commission are others. And nongovernmental research has
produced large amounts of data, particularly in biomedicine, that is now being shared openly to accelerate the pace of scientific discovery.

While Open Data is related to Big Data on one hand, it’s also related to the Open Government movement on the other. Open Government includes collaborative strategies to engage citizens in governing as well as the government releasing Open Data to the public. This book’s Appendix B, “Defining Data Categories,” gives a more detailed analysis of how Big Data, Open Government, and Open Data are related, complete with a Venn diagram.

The Open Business Opportunities

Although there’s widespread agreement that both Big Data and Open Data will be important business resources, no one is sure exactly what they’ll be worth. Determining the overall value of Open Data is far from easy. Many companies that use it are so new that it’s too early to measure their success. On the other hand, many established companies use open government data as just one resource for their work, making it hard to figure out how much it contributes to their business.

The Open Data 500 study, which I’m now directing at the GovLab at New York University, will give economists and other researchers a new information base to help assess Open Data’s value. This study, which is funded by the Knight Foundation, is the first real-world, comprehensive study of American companies that use government Open Data in health, finance, education, energy, and other sectors. We’re identifying 500 of these companies and surveying them to see how they use government Open Data and how they think government agencies can make their data more useful. We plan to make our findings available on a website by early 2014 where researchers can download our data, new companies can complete our survey, and members of the Open Data community can suggest future research.

To identify different kinds of Open Data companies, my colleagues and I began by looking at other research that had already been done.
Since 2012, the Open Data Institute in London has been working with the consulting firm Deloitte to study Open Data’s potential. In a series of studies led by Harvey Lewis, a research director in Deloitte’s Insight Team, the firm has identified five Open Data business “archetypes”:

- **Suppliers** publish their data as Open Data that can be easily used. While they don’t charge for the data—if they did, it wouldn’t be Open Data—they increase customer loyalty and enhance their reputations by releasing it.
- **Aggregators** collect Open Data, analyze it, and charge for their insights or make money from the data in other ways.
- **Developers** “design, build, and sell web-based, tablet, or smartphone applications” using Open Data as a free resource.
- **Enrichers** are “typically large, established businesses” that use Open Data to “enhance their existing products and services,” for example by using demographic data to understand their customers better.
- **Enablers** charge companies to make it easier for them to use Open Data.

I’ve found these categories useful and have also come up with two simple categories of my own.

The first I’d call Better Business Through Open Data. Open Data can improve healthcare, energy, education, finance, transportation, and many other aspects of consumer society. (In an interesting insight, the Deloitte team has noted that many sectors “will benefit most from open government data that has direct relevance to consumers, and stimulating interest from consumer-driven businesses may yield the greatest economic impact.”) In many of these areas, Open Data may become such a seamless part of daily life that most people won’t notice it’s there. Patients will still go to doctors, but now they’ll get better, more informed care. Households will still use electricity, but now they’ll use Open Data to help manage their energy use. Consumers
will use data-powered websites to choose credit cards and financial services as easily as they now use the web to book a flight. All kinds of daily activities will seem largely the same, but better, as consumer-focused companies improve their services by using Open Data behind the scenes.

In contrast, there are other opportunities that I’d call Open Data Pure Plays—companies, even whole industries, that simply would not exist without Open Data. They include startups that are revolutionizing agriculture by analyzing weather data; companies using Open Data to predict trends in healthcare, financial markets, or other fields; companies that manage and market government data; and companies developing market insights with data from the vast universe of social media.

Like the web itself, Open Data will be a major driver for new businesses and economic growth in the United States, the United Kingdom, and other countries. It will create new jobs, fuel new startups, and launch new industries with revenue in the billions. This book covers the full range of Open Data opportunities and the factors that will shape how Open Data is used in the years ahead.

About This Book: Four Open Data Promises

This book describes the business applications of Open Data with examples from dozens of companies, many of which I’ve interviewed myself. Part 1, which begins with this chapter, describes the four big Open Data promises for businesses and their clients: the potential for hot new startups, tools and services for smart investing, strategies for savvy marketing, and new forms of fast innovation.

Chapters 2 through 4 describe the startups that are rapidly developing by using Open Data.

Chapter 2 describes companies that are turning government data into dollars using the huge data resources that the U.S. government and others now make available for free. This chapter profiles a variety
of companies using GPS data in new ways; the boom in the use of government health data; the startup “incubator” at London’s Open Data Institute; and The Climate Corporation, recently bought for about $1 billion, which is working to revolutionize agriculture and increase farmers’ profits worldwide.

Chapter 3 covers choice engines for smart disclosure, a category of new companies that help consumers make informed choices in complex areas like healthcare, financial services, and transportation. Several well-known companies, from Kayak to Zillow, are already providing Smart Disclosure, but there’s room for growth here, particularly in the United States. In the United Kingdom, where several large companies provide Smart Disclosure on energy providers, credit cards, and insurance, an estimated 24 million people use these price-comparison websites every month. Based on population, that’s the proportional equivalent of 100 million monthly customers in the United States, a huge potential market.

Chapter 4 describes new companies whose goal is to “manage the data deluge” and profiles several new and old companies that help others work with government Open Data. While several large companies have provided government data services for years, the release of more Open Data is creating new opportunities and new startups. One of them, Enigma.io, recently won the prize as the hottest new startup at TechCrunch Disrupt NY, beating out about 30 other contestants. That’s the tech world’s equivalent of winning American Idol and is a sign of the potential of Open Data.

Chapters 5 and 6 describe how Open Data is creating new companies and tools for smart investing, providing business opportunities for both the entrepreneurs who create these services and the investors who use them.

Chapter 5, on data-driven investing, shows how Open Data and new analytic techniques are giving investors valuable tools and insights. A company called Capital Cube has developed software to analyze more than 40,000 globally traded public companies, update
its data on them every day, turn the data automatically into narrative descriptions of current trading, and provide tools for comparing prospective risk and return of different companies in different sectors. In the United Kingdom, Duedil (for “due diligence”) provides data on small- to medium-sized companies in the hope of encouraging hundreds of billions of dollars in new investment. With the Securities and Exchange Commission now adopting XBRL, a computer language that can encode large amounts of information, investors will soon have even more Open Data on public companies in a highly usable form.

Chapter 6 covers green investing, driven by Open Data, which has rapidly gone from being an idealistic goal to a mainstream investment strategy. Investors are increasingly looking for companies that operate sustainably and have the data to prove that they do. Sustainable companies aren’t just good for society and the environment—they tend to be models of good corporate governance and low operational risk. A large majority of S&P 500 companies now use sustainability measures from the Global Reporting Initiative to report on their practices. Bloomberg has begun offering sustainability data to its subscribers, and the number who use it is growing by almost 50 percent a year.

Chapters 7 and 8 describe how Open Data provides new opportunities for savvy marketing.

Chapter 7, on how reputational data defines your brand, covers a powerful form of Open Data that’s created by ordinary people. The complaints, experiences, and opinions that people post online add up to a body of reputational data that’s open for the world to see. Smart companies are now learning how to mine this data, understand it, respond to it, and use it to build their brands through sophisticated online marketing. The founders of PublikDemand, a new company that channels consumer complaints, describe how companies can turn complaints to their advantage: they’re championing a new approach
called “social customer service” that companies can use to improve their reputations by responding openly to public complaints.

Chapter 8 covers the marketing science of sentiment analysis, a cutting-edge technology for turning thousands of consumer opinions into quantitative business intelligence. Countless websites now invite consumers to write reviews of hotels, restaurants, or just about any other kind of product or service, and untold numbers of consumers tweet or blog about these products or services on their own. It can all be synthesized and analyzed as Open Data, creating datasets out of the collective consciousness of everyone who uses the web. Sentiment analysis extracts data from social media with particular attention to words that have a positive or negative “sentiment” attached to them. In this chapter, Seth Grimes, a guru of sentiment analysis, describes how companies can use this technology for new kinds of business and market intelligence.

Chapters 9 and 10 explore the fourth promise of Open Data—the way it can be used for fast innovation.

Chapter 9 shows how new forms of crowdsourcing can help research-focused companies find solutions using Open Data. It describes academically based “citizen science” projects that have recruited close to a million volunteers to analyze and refine different kinds of scientific data, providing a possible model for the private sector. And it examines companies such as TopCoder, Kaggle, and InnoCentive that use prizes and competition to recruit expert help for tough data problems.

Finally, Chapter 10 explores how Open Data is fueling open innovation—a dramatically new scientific approach that shares data early and openly to foster collaboration across many research labs. By sharing data that would normally be kept secret, open innovation runs counter to the prevailing cultures of scientific research and drug development. But it’s so effective that it may justify a new business model. Companies ranging from biomedical startups to the pharma giant GlaxoSmithKline are now using open innovation productively.
Looking Ahead: The Changing Business Environment

The 10 chapters in Part 1 show that Open Data’s potential is immediate and real and include pragmatic advice for any company that wants to apply it. But companies based on Open Data also have to watch the social, legal, and political factors that will shape how we use data in the years ahead. Part 2, “The Business Environment: New Trends in Open Data,” includes three chapters and a conclusion that describe the most important trends to watch.

Chapter 11 explains how privacy concerns, heightened by National Security Agency revelations, may change the rules around personal data of all kinds. Personal data can become Open Data in a special sense: with new technology, individuals can choose to make their personal data “open” for certain purposes, securely and selectively. Companies such as Personal.com and Reputation.com are betting that a new market in personal data, controlled by individuals, will open up new business opportunities while protecting personal privacy.

Chapter 12 shows how financial regulators, advocacy groups, journalists, and others are using Open Data to make business and government operations more transparent. It describes the new rules that successful businesses will have to follow to do well in this new “see-through” society. And it shows how some data-driven organizations are making transparency their business.

Chapter 13 describes the growing push for Open Data by national governments in the United States, the United Kingdom, and Europe, by state and local governments, and in other parts of the world. The United States has instituted an Open Data Policy that may have a historic impact, and the U.K. government, the European Union, and the G8 have made similar commitments to Open Data. These changes all promise that government Open Data will be a growing resource that businesses around the world can use for innovation and growth.

Finally, Chapter 14 summarizes the most important lessons of this book and gives a preview of the years ahead.
While Open Data has many sources and many uses, the Open Data movement began with a push to make government data more accessible, meaningful, and usable. Today, federal data is a huge, free resource that’s just beginning to be tapped. The next chapter explores the potential of government Open Data and the companies that are starting to put it to use.
About the Author

Joel Gurin brings a unique background to his work on Open Data. He began his career as an award-winning science journalist, became a leader and innovator in magazine and online publishing, led a White House task force on consumer-focused data and information, and now helps lead a foundation-funded program on Open Data and Open Government.

Gurin began work as a science writer and editor immediately after graduating from Harvard University with a degree in biochemical sciences. He has won the top science-writing awards from the American Association for the Advancement of Science and the National Association of Science Writers. He is the coauthor with William Bennett of The Dieter’s Dilemma, which introduced the setpoint theory of weight control, and edited the book Mind/Body Medicine with Daniel Goleman, author of Emotional Intelligence. Gurin was the editor and cofounder of American Health magazine, the first health publication to win the National Magazine Award for General Excellence.
Gurin had a long and successful career at Consumers Union, the nonprofit publisher of *Consumer Reports*, where he began as science editor, soon became editorial director, and then served as executive vice president for almost a decade. As EVP of Consumers Union, Gurin launched and grew *Consumer Reports'* website, ConsumerReports.org. Under his leadership, it became the world’s largest information-based paid-subscription site: it now has more than three million active paid subscribers.

In December 2009, Gurin joined the Obama administration as Chief of the Consumer and Governmental Affairs Bureau of the Federal Communications Commission. He conceptualized and served as Chair of the White House Task Force on Smart Disclosure, which studied how government Open Data can help consumers choose their best options in services like healthcare, financial services, education, and energy. He is now senior advisor to the Governance Lab at New York University, which studies how to use Open Data and citizen engagement to improve the ways we govern and thus improve people’s lives. He lives in Scarsdale, New York, with his wife, Carol; they have three grown children.