AUTHORS ZACH GEMIGNANI • CHRIS GEMIGNANI
FOUNDERS OF JUICE ANALYTICS

DATA FLUENCY

EMPOWERING YOUR ORGANIZATION WITH EFFECTIVE DATA COMMUNICATION

Ex. 02

WITH DR. RICHARD GALENTINO
& DR. PATRICK SCHUERMANN

WILEY

FLOWINGDATA.COM

Data Fluency

Data Fluency

Empowering Your Organization with Effective Data Communication

Zach Gemignani

Chris Gemignani

Dr. Richard Galentino

Dr. Patrick Schuermann

WILEY

Data Fluency: Empowering Your Organization with Effective Data Communication

Published by
John Wiley & Sons, Inc.
10475 Crosspoint Boulevard
Indianapolis, IN 46256
www.wiley.com

Published simultaneously in Canada

ISBN: 978-1-118-85101-2
ISBN: 978-1-118-85089-3 (ebk)
ISBN: 978-1-118-85100-5 (ebk)

Manufactured in the United States of America

10 9 8 7 6 5 4 3 2 1

Library of Congress Control Number: 2014946679

To our parents, who shared a love of art and joy of teaching
that we try to pass on to those communicating with data.

About the Authors

This book was a collaborative effort built upon years of experience helping companies make better use of data. Zach led the writing effort and defined the Data Fluency Framework that is the foundation of this book. Chris is responsible for many of the design and data visualization ideas and approaches that we share. Richard contributed from his experience in healthcare, education, and nonprofits, conceived of the Data Fluency Inventory, and took on the task of coordinating with our editors. Patrick worked with our research associate Tim to contribute content on organizational development, helping make this book a tool for leaders interested in transforming their organizations.

Zach Gemignani is co-founder of Juice Analytics and has helped build the company's reputation for designing engaging information experiences and delivering unique data visualization solutions. As CEO, he is responsible for the strategic direction, thought leadership, and business development of the company. Prior to Juice, Zach led reporting and analytics efforts at AOL and was a consultant with Diamond Technology Partners and Booz Allen, where he developed a reputation for creating exquisite slide presentations. He graduated from Haverford College with a Bachelor of Arts degree in Economics and received his MBA degree from The Darden School at the University of Virginia. Zach lives in Nashville, TN with his wife and three children.

Chris Gemignani is co-founder of Juice Analytics and the company's technology visionary. Chris earned his data chops in the credit card industry, taking on responsibility for risk modeling and analyzing cardholder behavior patterns. He combines this analytical experience with the ability to bring these insights to the screen with a hypercritical eye for user interface and interaction design. Chris graduated from Williams College with a Bachelor of Arts degree in Computer Science and Economics. He received a Masters in Economics degree from Washington University in St. Louis.

Dr. Richard Galentino. serves as CEO of Stratable, Inc., a strategic planning and organizational development consulting firm. Prior to launching Stratable, Richard led an international medical effort sending hundreds of doctors, nurses, and allied health professionals to more than 27 countries. Selected as a Harvard International Education Policy Fellow, Richard is a graduate of Harvard University (Administration, Planning, and Social Policy; Ed.M.) and Georgetown University's School of Foreign Service (Economics; B.S.F.S). Richard earned his doctorate in education leadership and public policy from Vanderbilt University. Richard and his family reside in Nashville, TN.

Dr. Patrick Schuermann is a research professor at Vanderbilt University's Peabody College of Education. Having previously served as the director of policy and technical assistance for the national center on educator compensation reform for the U.S. Department of Education and the PI for numerous research projects in school leadership and education technology, Patrick currently serves as the director of the independent school leadership master's degree program and chair of the Peabody professional institutes. Patrick resides in Nashville with his talented singer-songwriter wife and their two dogs.

Credits

Executive Editor
Carol Long

Project Editor
Adaobi Obi Tulton

Technical Editor
Nathan Yau

Production Editor
Christine Mugnolo

Copy Editor
San Dee Phillips

Manager of Content Development and Assembly
Mary Beth Wakefield

Director of Community Marketing
David Mayhew

Marketing Manager
Carrie Sherrill

Business Manager
Amy Knies

Vice President and Executive Group Publisher
Richard Swadley

Associate Publisher
Jim Minatel

Project Coordinator, Cover
Patrick Redmond

Compositor
Maureen Forys, Happenstance Type-O-Rama

Proofreader
Nancy Carrasco

Indexer
Robert Swanson

Cover Designer
Wiley

Cover Image
Courtesy of Zach Gemignani and Chris Gemignani

Acknowledgments

A decade ago, Chris and I set out to create a company that bridged the gap between data and the people who might use it. This book draws from the many lessons we've learned from our discussions with colleagues and clients about presenting, visualizing, and sharing data.

I'm grateful and humbled by the energy and commitment demonstrated by all my colleagues at Juice. A special thanks to Ken Hilburn, James Lytle, and Michel Guillet for helping stretch our thinking along the way. Coming to work is always a joy when I get to collaborate with talented individuals including Djam Saidmuradov, Meghna Kukreja, Lindsay Conchar, Tim O'Guin, and Jennie Gemignani.

I'd like to thank Nathan Yau for inviting us into the Wiley family. His unflagging dedication to learning, sharing, and discussing all elements of data visualization is impressive and inspiring. As our project editor, Adaobi Obi Tulton has been a patient guide through the process.

I also appreciate the thoughtful efforts of Tim Drake, who contributed to the writing and research of the book. Tim is a Ph.D. student in Education Leadership and Policy at Vanderbilt University. Tim writes, researches, and teaches in the areas of quantitative research design and methods, data-driven decision making, and K–12 education leadership and policy.

Finally, a heartfelt thanks to my wife, Andrea, and kids—Owen, Maya, and Lila—who have been supportive and patient as I took on yet another responsibility.

Zach

Contents

Foreword

It's been more than a decade since I took my first statistics course in college. Unlike for many, my introduction to statistics brings back happy memories of an enthusiastic professor who jaunted up and down the stairs of the lecture hall. It's not easy to get excited about beginning concepts in distributions and hypothesis testing, but he pulled it off. I grew interested in working with and understanding data which eventually led to many years of graduate school. I had no clue back in college that statistics—or more generally, using data—would be so popular now. I just liked to play with data. And there's a lot of data to play with these days.

Every day I read or hear about companies and organizations that use data in some way. There's a wide array of applications: improving business, providing better service to customers, helping to make the lives of others easier, or communicating complex processes. There's an excitement. People want to gain insights from all this data they collected.

There's a gotcha though, and it's a big one. You can't just take a stream of data, plug it into the most expensive software you can find, and gather instant results—regardless of whether you're one person or a big organization. It's never that easy. Anyone who tells you otherwise either doesn't know what he is talking about or is trying to sell you something.

As someone focused on data visualization, I would love to build a dashboard or develop an interactive tool that enables people to understand their data in an instant. No background needed. However, you have to learn how to use the tool before anything worthwhile comes of it. You must know what data represents and how to analyze and interpret.

When you start to look at how an entire organization can grow more fluent in the language of data, you introduce other challenges. Those in management have different responsibilities than those working on the floor, but there must be a proper foundation for everyone to work together in an effective way.

Zach and Chris Gemignani, co-founders of Juice Analytics, help groups with these challenges every day, and now they educate others with *Data Fluency*. The two brothers and their team

have been consulting long before "big data" became a thing, before Google's chief economist Hal Varian said that the job of a statistician is sexy, and before I started FlowingData. The Gemignanis' experience shows in their articles online and in this book. Their advice is practical but general enough so that you can apply frameworks to your own situation.

When I first searched for "data visualization" years ago, the Juice Analytics site was one of the first ones I found and still subscribe to today. So I was excited when Zach and Chris agreed to write *Data Fluency*. However, this isn't a book about visualization. It certainly covers the topic, but *Data Fluency* provides a wider view.

When you have visualization floating around in your organization—reports, talking slides, and data displays—does it actually matter if no one looks or gets anything out of it? It ends up in the recycle bin or as background noise. You can have the most efficiently designed charts in the world, but at the end of the day, you need people to pay attention. The goal is to bring data closer to the front so that everyone from management on down can make better informed decisions.

At the same time, there is no promise of a panacea or a new tool to make all data problems go away. It's a realistic view that stems from the Gemignanis' experience. They understand that often a lot of moving parts in groups might move slower than others or are difficult to change. I'm just a one-man show with FlowingData, but in my own consulting work, I understand the pains of bureaucracy all too well. The key is to work with the areas that do change and go from there. *Data Fluency* is an excellent guide to figuring out how you can do this.

Sitting here, thinking about what data might look like another decade from now, I can only imagine more of it, at a more detailed level. In the present day, the rate of collection far exceeds the rate at which we can understand. However, the growing rate at which people *want* to understand is a different story. So the more people who can learn the language of data now, the better we will be for it later.

Nathan Yau

Introduction

How do you change minds?

My brother and I huddled in my basement, putting the finishing touches on our analysis. The sun had set and our presentation was the following morning. We had spent the last month gathering data about student retention at online schools. We wanted to know what caused students to leave and what kinds of students tended to stay.

We had the slides to share with the executive team. The presentation summarized an attrition model, segmented the student population, and offered recommendations. Yet we felt something was missing.

How could we teach people to care?

Behind our numbers were individual students who chose the online school, took out student loans to pay for their education, spent hours with the online coursework, consulted with teachers, and tried to keep up with the schedule. Our analysis flattened the individual stories, successes, and struggles of these students. How could we bring real life back into our presentation?

As the clock ticked toward midnight, we got an idea: We'd create an animated movie. It would show how every student found their way into the school from different points of origin, how they progressed through their schoolwork, and how they eventually made the decision to stay or leave.

The movie-making was more quick and dirty than elegant. We constructed images showing where each student existed on their journey then joined them into a single image for each day of the school year (Figure 1). The students were positioned precisely and moved like stop-motion figures. Finally, Chris wrote a script to generate hundreds of single-day snapshots then weaved

them together into an animation. For a couple of tired data junkies, our couple minutes of movie magic felt like a Spielbergian masterpiece.

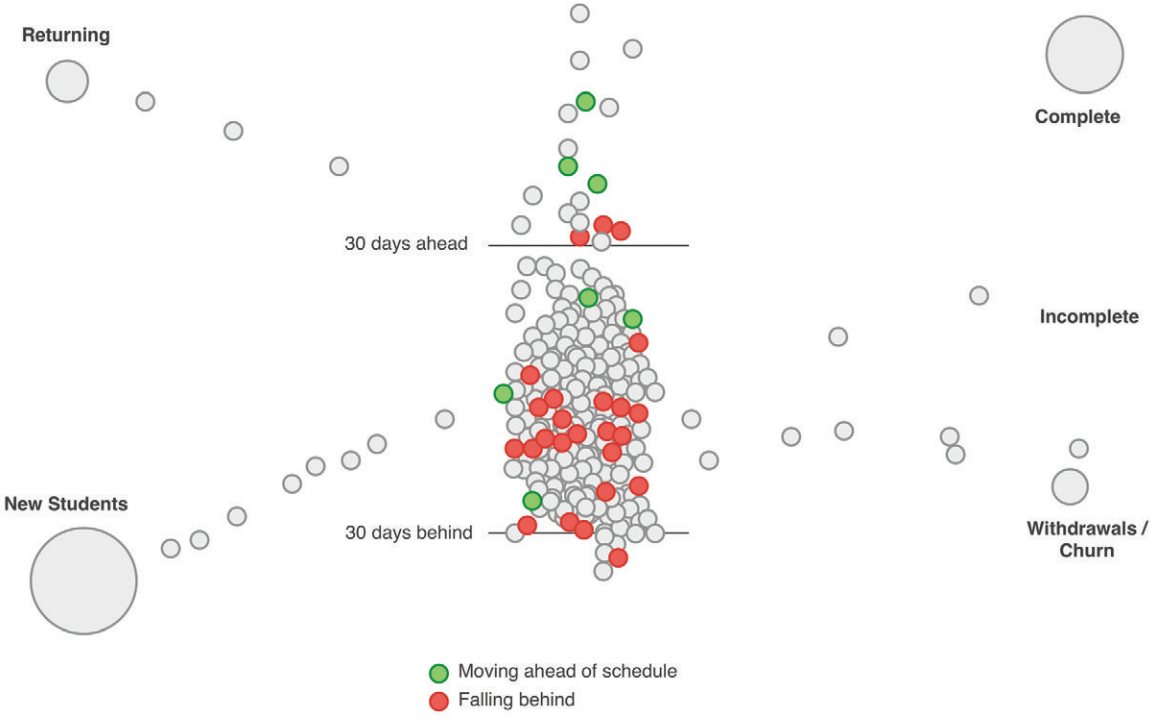

FIGURE 1 *A point in time from our movie about student retention*

The students marched across the screen on their way to joyful completion or disappointing withdrawal. The data had new life. And it sparked a conversation with our client.

That creative exercise ignited a passion. We had started Juice Analytics a few months prior, knowing that we wanted to help businesses gain insight and understanding from their data.

That night helped us turn from focusing solely on the numbers to how they are communicated. We realized we wanted to find better and more creative ways to help people understand data. Could we bridge the gap between data analysts and the people who can take action from their work?

For almost a decade we have pursued this goal. Juice Analytics has worked with over a hundred companies—from start-ups trying to deliver data to their customers to global brands looking for better ways to communicate data to executives. We've designed engaging interactive dashboards, reports, and analytical tools—all with the goal of helping real people make sense of and act on data.

Along the way, we've learned a few important lessons.

DATA IS THE NEW LANGUAGE OF BUSINESS

Data is a medium to communicate and convince. Its value has been recognized and elevated over the last few years. Media sites such as FiveThirtyEight (from ESPN) and Upshot (from *The New York Times*) are creating public discussions that combine data analysis and visualization with journalistic storytelling. These sites are a public expression of a trend that has been percolating within many smart organizations.

However, not everyone is comfortable communicating with data. Many of the audiences we design for—administrators, attorneys, marketers—are unfamiliar and inexperienced with getting value from data, even in small doses. One of the great challenges of data communication is building a dialogue. As much as a speaker must express himself through clear, accurate data presentation, the audience needs to be a willing and capable recipient. Presenters of data need to meet their audiences where they are, in ways that their audience can comfortably engage with the content.

How do you create common ground for more effective data communication? You can start by teaching the fundamental grammar of data visualization: metrics, dimensions, distributions, relationships, outliers, and variance. You can encourage good choices for how to express data by picking the right chart to emphasize the important elements in the data. You can learn from the expert data communicators to see how they fluently use the language of data.

More than ever, data will be a large part of how you convey messages. You need to ensure that everyone in your organization can participate in the discussion.

DATA COMMUNICATION IS A SOCIAL PROBLEM, NOT A TECHNOLOGY PROBLEM

For years, many organizations found it important to strive for data volume and invest in bigger databases and feature-rich platforms. Lost in the focus on size was the real prize—actionable insights in the hands of people who can do something about it. The first generation of business intelligence was about delivering complex, full-featured solutions designed for the IT team. Yet vast quantities of data collected by organizations remain disconnected from the people who might make use of it.

Making data useful is a problem that ultimately must be solved by people—people who understand the specific context of the data, people on the front-lines of decisions, and people who deeply understand the problems that data can illuminate. Data is useful when people use it to tell stories, craft compelling visualizations, and construct thoughtful analyses. People are the missing ingredient.

Unlocking the value of data takes more than individual efforts. It takes the interactions between people who communicate with data, discuss meanings, and debate what actions to take. There is a need for a data culture within organizations that embraces informed decision making.

Data is a cold, lonely medium on its own. Data needs to be humanized and human-sized. It needs to be made relevant to the audience by being clearly linked to relatable problems. It should be presented in intuitive, visual, and simple ways. And like any language, data should be about conveying meaning.

CONNECTING AND COLLABORATING

Much of the conversation on data occurs across the great divide between those who have a cultivated knowledge of data and those who have responsibilities that seldom involve digging into data. There are language barriers, biases, and misconceptions between these groups of people.

On the one hand, consider a data analyst who has created a complex spreadsheet that helps explain inefficiencies in operations or perhaps defines

marketing channel attribution. To do this they have learned the intricacies of APIs, how data is collected and defined, where it is gathered and how it can be joined to other data. To them, data is a flashlight illuminating a bit of truth in a chaotic world.

The analyst's boss comes from an entirely different world. She is equally smart and invested in finding ways to improve the organization's bottom line, but she has little attention or time for a detailed spreadsheet or black-box data algorithm. She's more likely to be moved by a compelling story than a table of numbers.

If these types of people can find a way to collaborate, the organization can benefit. The analyst's work can see the light of day and drive smarter decisions. His manager can help keep the focus on the pressing problems where better analysis can impact actions.

Our goal has been to create a productive dialogue among those who are data fluent and those who are just learning the language of data. If we can do this, we can connect people who can ask the best questions with those who can answer the questions.

TURNING DATA INTO ACTION

I was presenting to an audience about the untapped value of data when I saw a hand shoot up in the back of the room. It belonged to a serial entrepreneur who I had known for years. He commented to the crowd, "Data isn't valuable. In fact, it is costly. Think about all the money that goes into gathering, storage, management, and software. The insights that can be found in the data, isn't that what's valuable?" He offered a valuable distinction. But maybe we should go further. It is only through actions taken that true value that can be unleashed from data.

The data industrial complex—big business intelligence providers like IBM, MicroStrategy, and SAP—have plenty of incentive to make you believe that gathering more data is a source of value. More data needs more powerful tools with complex feature sets. Bigger data is better data.

Don't buy it.

A core principle of product design, according to Joshua Porter, Director of User Experience for HubSpot, is "Usefulness is job #1." He goes on to say, "If your product is not *useful*, if humans do not find use for it, then the design has failed. Your product must help people do something valuable in their lives."[1]

Your data is made useful when it helps people do something better in their job. When data is communicated in ways that are easy for people to grasp, the information can drive dialogue and discussion. The insights and stories in data can get people talking in a productive way. And from these discussions come better decisions and actions. Without this journey of insight to action, all your data might as well be hot air.

VISUALIZATION IS ONLY A PIECE OF THE SOLUTION

The human brain has an incredible ability to absorb and process visual information. We may not remember someone's phone number or the name of that person we just met, but our visual systems can put computers to shame. Those annoying CAPTCHA systems used to verify that you aren't a spamming robot are evidence that our minds can process information and find patterns with ease. This is the foundation on which data visualization has been built.

Over the last decade, we have had a front-row seat to watch the explosive growth of data visualization. In 2005, we started blogging about data visualization and a handful of passionate practitioners and academics were at the core of a fledgling data visualization community. Today, there are dozens of conferences, established academic programs, and thousands of designers eager to visualize your data in an infographic.

Lost in this beehive of activity is a simple fact: Data visualization is but a means to an end. That end is to effectively communicate ideas and insights by transforming and representing numbers in ways that everyone can understand. The means can—and must—go beyond charts, infographics, and sophisticated interactive visualizations.

How do we reach beyond infographics? Helping an audience work with data requires creating a logical flow through the information—a flow that has a clear beginning and an end that can lead to actions. Good data communication means providing guidance about the meaning of data elements and timely

insertion of contextual information. Presenting data isn't just about how that data is visualized, but also about how the user can interact, explore, and have a great experience that works with what they already know.

If data is a new medium for communication, there is something to be learned from the many other forms of communication that have come before—like print, photography, and film. A movie director does much more than string together a series of images. He connects to the viewer by artfully combining elements like music, sound effects, editing, and cinematography.

The role of the data communicator is similarly complex. The goal should be to create "information experiences" that transform how audiences think about a subject and make better decisions.

Whether better decisions are made with data often depends on organizational dynamics. Organizational culture deeply influences whether "data products" lead to better decisions. Processes need to support effective data communication. And, in dynamic organizations, effective data communication constantly shapes better processes, systems, and decisions.

It is these lessons, and many more, that we share in *Data Fluency*. Our ambitions are broad. We hope to construct a roadmap that organizations large and small can use to improve how they work with data.

WHO IS THIS BOOK FOR?

The journey to data fluency is important for any organization that wants to have data inform decision-making. It takes a diverse set of people to build the skills and culture for data fluency. Whether you are a leader or an analyst, this book offers practical guidance to help you and your organization on this journey. Data fluency requires:

- **Organizational leaders** who want to create a culture of data communication and usage, and need to build a team of savvy data consumers and analysts.
- **Data novices** who are beginning to learn how to communicate with data.
- **Experienced analysts** who work deeply with data and need to effectively communicate their findings.

- **Data product authors** who design and create dashboards, reports, and data-rich applications, and want their work to engage their audiences.
- **Technology leaders** who are responsible for managing data, creating data solutions, and ensuring the organization extracts maximum value from its data assets.
- **Human Resource managers** who need to select and train their staff with the skills necessary to work with data.

The first chapter explains the Data Fluency Framework in detail. With an understanding of the Data Fluency Framework, Figure 2 identifies a few of the best chapters to focus on if you are interested in quickly getting to the content most relevant to your role.

Chapter	Organization Leader	Technology Leader	HR Manager	Analyst	Content Creator	Data Novice
3. **How Organizations Struggle** Diagnose the challenges your organization is facing	✓	✓				
4. **Consumers Guide to Data** The skills you need to have to become an effictive consuer of data			✓			✓
5. **Producers Guide to Data** The skills to communicate with data		✓	✓	✓	✓	✓
6. **Data Fluent Culture** Creating conventions and standards for a data fluent culture	✓		✓	✓	✓	
7. **Data Product Ecosystem** Creating an environment for the fluid and efficient sharing of data presentations	✓	✓			✓	

FIGURE 2 *Chapter Guide*

NOTE

1. Porter, Joshua. "Principles of Product Design." *Bokardo.com*, n.d. Retrieved from `http://bokardo.com/principles-of-product-design/`

CHAPTER 1

The Last Mile Problem

Water, water, everywhere, and all the boards did shrink;
Water, water, everywhere, but not a drop to drink.

Samuel Taylor Coleridge
The Rime of the Ancient Mariner

The Texas Oil Boom at the turn of the 20th century saw populations in Dallas double and double again in a few short years. This tremendous growth was driven by a wholesale shift of the U.S. economy and infrastructure to automobiles. The first Ford Model T left the factory on September 27, 1908. Less than 20 years later, in 1927, Ford had produced 15 million cars as supply and demand caused a massive societal shift—the United States literally drove into a new society.

Similarly, the advent and widespread use of the personal computer at the end of the 20th century led to a new boom: the Information Age. Thirty years after the launch of the personal computer, we entered a new era of big data and data-driven decisions that has been compared to the oil boom. Certainly the hype is gushing and hopes are high that *data will deliver smart insights and a more intelligent enterprise.*

This is the promise, but here's the hitch: Although generations of teenagers have learned how to drive as a rite of passage into adulthood, and with it have found new degrees of freedom, we have not systematized the acquisition of skills essential to ensure data fluency. The language at the heart of the highways of commerce, now, and even more so in the future, is available to relatively few.

The 1920s and 1930s saw the advent of factories to efficiently mass *produce* automobiles, and generations of insight inform our patterns of *consumption* of this critical product. Almost 100 years later, a critical set of skills is now needed to ensure individuals and organizations are thoughtful *consumers* of data, and emerging skill sets are essential to *produce* data-based presentations and actionable data products.

It is to the development of this critical set of skills—those of being informed, capable consumers of data, and of being accomplished producers of data presentations and products—that this book is dedicated. Our goal is to help individuals and organizations understand and develop data fluency, as we contend it is the new language, the new highway, of commerce in the 21st century.

THE INFORMATION AGE: DRIVING THE NEED FOR DATA FLUENCY

Fantastic advances in data storage capacity have fundamentally changed the trade-offs we need to make regarding what to keep and what to delete. Rather than having to carefully decide what elements of our digital reality to capture and which to throw away, we can now keep everything. We can have it all—and we do. According to independent research organization SINTEF (The Foundation for Scientific and Industrial Research), 90 percent of all the data in the world has been generated over the last two years.[1] With more Instagram pictures, more tweets, more history of where customers go on the web, we are rapidly growing the amount of data we can sift through.

In a sense, the raw materials for informed decision-making have never been more plentiful. Yet the promise of data nirvana still seems far off. Students, scholars, employees, and executives are often still making crude decisions based on chance, gut, or whims of the crowd. In this era of data as the new oil boom, where's the payoff? Are we making better decisions and are we better able to understand our world? Are we driving cars or still riding horses along the digital highway?

On the ground, in the organizations we've worked with at Juice Analytics, people are often frustrated by their inability to effectively use data. They've built data warehouses, invested in expensive business intelligence solutions, and spent finite fiscal resources to hire data scientists. They've data-mined, analyzed, defined key metrics, and created dashboards. Despite these efforts, data is often under-used and misunderstood.

Few people, and fewer organizations, consistently engage with the data and use it to guide their thinking. Our vision is for everyone, from front-line customer service agents to senior executives, to leverage the mountain of data at their disposal. Forget the complex Wall Street trading models or IBM's Watson computer diagnosing disease—data in your organization can and should be used in simple, incremental ways to improve conversations, focus resources on priorities, and make small, everyday decisions with clarity.

Making use of data is a problem common to organizations large and small, public and private, and across market segments. According to a study conducted by the consulting firm Avanade, "more than 60 percent of respondents said their employees need to develop new skills to translate big data into insights and business value."[2]

With all the promise that data holds, and the hope that data can help us make more informed decisions, the big question is: What is causing the gap between the vast opportunity of data and the reality of organizations struggling to act on this data? Here are a few theories:

1. Many people are data phobic and unwilling to engage with data to make decisions.

While at a leading Internet media company, we witnessed analysis teams dutifully churning out detailed reports about how online content was performing while the report recipients, content managers, dutifully ignored the information. Decisions had always been made based on gut and continued despite more and more detailed data about content usage and users.

2. Technology and personnel limitations constrain organizations' ability to work with their data sources.

Many organizations we encounter lament their spreadsheet-driven culture. Every department has its own mechanism for gathering, analyzing, and reporting on its unique data. No consistent "source of truth" exists and data analysts become indispensable because they are the only people in the organization who know how a financial model works, how to access and understand the data sources, and its strengths and weaknesses. People in these organizations wish for a technology solution that could bring all the information together and make it available to all decision makers in interactive, visual dashboards.

3. Organizational constraints inhibit the effective use of data.

In school districts around the country, superintendents often lament the lack of good data. Indeed, the Gates Foundation is currently putting significant resources into developing district level dashboards to inform decisions. Yet, critical organizational challenges remain with respect to collecting the data in a timely manner, linking data to competency and performance assessments, and engaging teachers in the process.

We believe that data-phobia, technology limitations, or organizational dysfunction are symptoms of something broader—not the root causes of the lack of payoff we are currently realizing from data. The root cause is something we call "the last mile" problem. Fundamentally, failing to use data isn't a technological problem, but a social problem.

The last mile analogy comes from telecommunications where bridging the final few feet from the big pipes carrying gigabytes of Internet traffic throughout your city to each individual house is the most costly. With data, collecting and storing information is the easy part. The technologists have done their job. It is analytics, application, and adoption that pose the greatest challenge. Although data storage can be done *en masse*, the last mile is personal and often organization-specific. Revealing insights, influencing decisions, and taking action requires skill and motivation at a personal and organizational level. This is the missing link—the last mile—requiring individual and organizational data fluency.

This book is about how organizations can more effectively communicate with data—both internally and with external constituents. It is about people and the specific skills needed to be capable consumers and effective producers of data-based reports and presentations.

DATA FLUENCY: UNLOCK THE POTENTIAL ENERGY OF DATA IN YOUR ORGANIZATION

In many ways, data is like oil—and it is certainly so in the economic engine of your organization. Just like you can't pull crude oil from the ground and pump it directly into your gas tank, or mold it into a plastic LEGO® brick, you can't dump data into an organization and expect it to be useful. Creating

value from data is a complex puzzle; one that few organizations have solved. Although there isn't a simple answer (and thus why so many organizations struggle), the good news is that understanding the nature of the problem offers a starting point for our path forward. Data fluency is the path—the ability to use the language of data to fluidly exchange and explore ideas that are important to your organization.

In this book, the goal is to help you unlock the potential of data in your organization. Your data challenges have less to do with technologies and organizational constraints, and more to do with developing the capacity of data consumption and production within individuals and organizational teams.

Data fluency applies to individuals (everyone needs the skills to "read and write" and "listen and speak" using data) and also to organizations that must create an environment that rewards productive data conversations. There are many practical resources and books that address individual data skills. However, there are fewer resources for helping transform an organization to achieve data fluency. This book draws on foundational organizational development literature as well as best practices from current industry leaders.

The goal is to offer a framework that can help you understand the pieces required to construct a data fluent organization. At the same time, it provides practical guidance that you can act on. You don't live in the theoretical, so the insights in this book won't stop there, but instead are rooted in real-life examples intended to provide actionable guidance.

The advice will be technology-agnostic. There are many tools that can help you communicate with data more effectively. You will like some more than others, and circumstances often dictate what you have to work with. The guidance offered in this book can be applied across the full spectrum of particular products or platforms you utilize and can help you make the best of them.

BIG DATA AND DATA METAPHORS

Data is everywhere. Peta- and exa- and zettabytes of data. Amounts that require mind-blowing analogies, like the 40 zettabytes of digital data that will be generated by 2020—a figure estimated to be 57 times the amount of all the grains of sand on the beaches of the earth![3] Though the amount is staggering, data is also captured from more sources and at a faster pace than ever before. In lockstep with the increase in the amount and kinds of data has been the technological capacity to house, store, and analyze it.

Big data—that is, the *volume*, *velocity*, and *variety* of information currently being gathered and stored[4]—represents the next great frontier, not only for business, but also for other sectors, including government, education, and healthcare. Anecdotes about the sophisticated use of data can catch your imagination and excitement:

> *Billy Beane's use of statistics to help the Oakland A's acquire undervalued players, an act that led to a playoff berth for a team that spent one-third the amount on salary as the New York Yankees.*
>
> *Google analyzing Internet search patterns to provide early warning signals for flu and dengue fever outbreaks.*
>
> *The United States Federal Government funding initiatives across departments to "greatly improve the tools and techniques needed to access, organize, and glean discoveries from huge volumes of digital data."*[5]

Technology professionals, business leaders and academics all view data as a means to forecasting, reduce costs, identify new business opportunities, improve research, and improve sales efficiency and effectiveness. As a result, big data startups have flooded the market. Venture capitalists are investing more than one billion dollars a year in the sector, funding hundreds of companies.[6]

The possibilities truly seem endless. Imagine, all your enterprise data, seamlessly integrated, available anywhere, anytime, just waiting to be translated into dollars. In this regard, data truly does seem like the new oil. In fact, there have been many metaphors used to describe our current era of big data. Tyler Bell, who is a big data industry analyst for O'Reilly, classified these according to the following typology:

- ***Natural resources*** *("the new oil," "goldrush" and of course "data mining"): Highlights the singular value inherent in data, tempered by the effort required to realize its potential.*
- ***Natural disasters*** *("data tornado," "data deluge," data tidal wave"): Frames data as a problem of near-biblical scale, with subtle undertones of assured disaster if proper and timely preparations are not considered.*
- ***Industrial devices*** *("data exhaust," "firehose," "Industrial Revolution"): A convenient grab-bag of terminologies that usually portrays data as a mechanism created and controlled by us, but one that will prove harmful if used incorrectly.*[7]

Of course, the metaphor that may resonate with you is largely determined by your organizational capacity. Are you like 73 percent of a recently surveyed group of U.S. business and IT executives who admit converting large data volumes into actionable intelligence remains a challenge?[8] Or are you like more than one-half of a representative sample of business leaders in the world that say that they have more data than they know what to do with? Do you view data as crude oil, something to be extracted, processed, and refined, a precious resource that must be converted if it is to be transformed into something of value? Or like gold, something to be targeted and mined? Or maybe a fire hose, essential when there's a fire, but overpowering when you just want a drink?

Clearly, embedded in these metaphors are a range of attitudes toward the data boom—from a fever pitch of excitement to dire warnings about the use (and abuse) of big data, warnings seemingly substantiated by high-profile data leaks.

Falling somewhere in between these two extremes are the vast majority of people, individuals who view data as valuable but are frustrated by an inability to capitalize on its value. Even leading research company Gartner sees big data slipping into the *trough of disillusionment* as interest begins to wane and organizations become frustrated as implementations fail to deliver value.[9]

So what has caused the fall? Why are many beginning to lose faith in the possibility and promise of data?

Simply stated, there is a growing recognition among many that gathering data, developing sophisticated storage systems, and hiring data scientists will not automatically translate into a competitive advantage. The efforts of countless executives, healthcare companies, school districts, and other governmental and nongovernmental organizations are a testament to this. A 2013 survey of companies that have implemented business intelligence solutions, shows that just 28 percent feel that the investment has delivered significant business impact.[10] The frustration with data, then, stems not merely from failed expectations, but from the inability of many end users to be thoughtful producers, consumers, and users of it. This book attempts to remedy this problem at both the individual and organizational level.

OUR DATA FLUENCY FRAMEWORK

With the goal of helping you unlock the potential of data for individual work, for collaborative working teams, and entire organizations, we have developed

a framework for data fluency. The framework, as shown in Figure 1-1, portrays the skill sets and competencies that you can develop.

The framework specifies two primary categories of skill required to develop data fluency, namely those required to be an expert *consumer* of data presentations as well as the expertise required to be a skilled *producer* of data presentations. The development and application of these skills occur at two levels: the *individual* and the *organizational*. The goal is to catalyze the development of data fluency associated with all four quadrants of the framework.

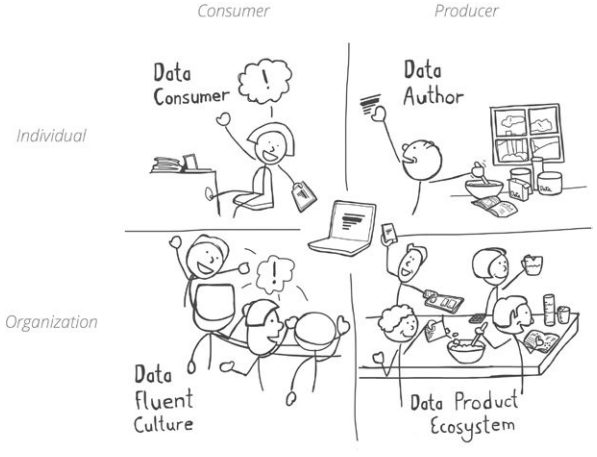

FIGURE 1-1 *The Data Fluency Framework*

CASE STUDIES: A WINDOW INTO THE FRAMEWORK FOR DATA FLUENCY

Subsequent chapters of this book are dedicated to helping individuals and organizations develop skill in each quadrant of the Framework for Data Fluency. This section provides four case studies to provide an initial window into the framework and illuminate the type of skills it addresses. For each struggle or promising practice outlined in the cases, subsequent chapters provide a set of actionable guidelines for individual and organizational growth.

Data Consumers: Fantasy Football

It may not be a stretch to say more Americans have learned about data and statistics through fantasy football than every college statistics course in the country. Each week, some 19 million NFL football fans spend their Sundays meticulously setting team line-ups based on statistical projections, historical patterns, and analysis of week-to-week variance. The couch potatoes who once relished on-field hits and in-game strategies now spend an average of more than eight hours a week diving into the data of the sport.

For the uninitiated, fantasy sports let fans play the role of team owners and managers by picking players for their own fantasy team and making weekly roster decisions. As the action plays out each week on the field, fantasy owners collect points against other competitors within their fantasy leagues. To

win, fantasy owners quickly realize that success often depends on studying player and team performance data closely.

Here are a few ways that NFL fantasy players incorporate data into their thinking:

Variation in Player Performance

The best fantasy owners understand the nature of week-to-week variance and its relationship to earning points. For example, touchdowns generally earn a fantasy owner six points; but touchdowns occur rarely and can fluctuate wildly. In contrast, the number of touches players receive may be a better indicator of how much the team is using them and their opportunity to provide the owner with points. Because *consistent* performance matters, successful owners often focus on players with more stable predictors of success (for example, touches) versus more sporadic events (for example, touchdowns).

Rankings Can Be Misleading

Fantasy football cheat-sheets offer rankings of players in every position. These rankings mask the differences and dispersion of expected performance. For instance, the top running back may be expected to perform 20 percent better than the second rated running back, who in turn is only expected to score five percent more points than the third through sixth rated running back. The data shows that players often cluster into tiers of performance. This statistical understanding was publicly explained by Boris Chen who stated that "players within a tier are largely equals. The amount of noise between the ranks within a tier and actual results is high enough that it is basically a dice roll in most situations." This concept has been widely adopted by fantasy owners as a player drafting strategy.

The Only Constant Is Change

The worst fantasy football owners are stuck in the past and pick players and teams that they have relied on in the past to generate points. That is, they fail to update their assumptions about the best teams, players, and trends. Following the data closely reveals when certain players have gone past their prime and when teams that once had high-scoring offenses can no longer put up big points. Clinging to past success may be a formula for disaster because the only constant in fantasy football is change.

Context Fills Out the Picture

Data viewed in isolation can be deceiving. Say, for example, that your top wide receiver scored only one-half the number of points that he scored on