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.
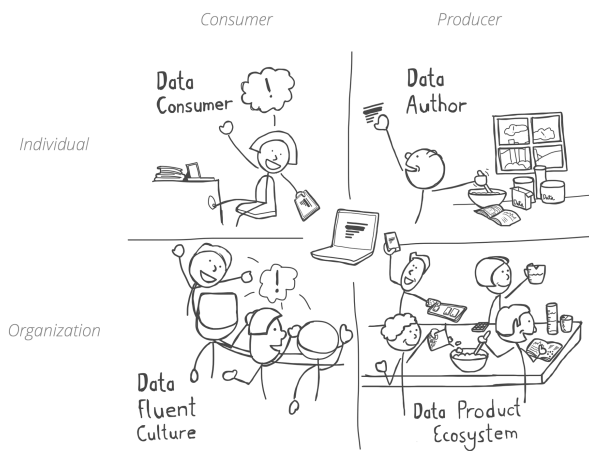

FIGURE 1-1 *The Data Fluency Framework*

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

average in a season. Is this a new and troubling trend? Should you trade? A little research might reveal that he matched up against one of the league's top cornerbacks, or his quarterback was knocked out of the game, or perhaps he tends to perform poorly in cold weather, away games. These environmental factors make a difference with respect to outcomes. Performance data cannot be understood in isolation—context matters.

So how did fantasy football create legions of fans who have developed a specialized dialect of data fluency? It has been a combination of education, effective data presentation, common data conventions, and incentives. Fantasy football owners have been taught how to use data to their advantage through the efforts of the NFL, ESPN, Yahoo!, and a cloud of other websites dedicated to football analyses. Organizations like Football Outsiders built new media businesses around data modeling and projections of player performance.

Leading online fantasy football sites like ESPN and Yahoo! have been aggressive in pushing data and data visualizations to their users. These sites include trend charts for every player, drive charts, player comparison graphics, and predictive models for estimating game outcomes.

The educated fantasy football community is also highly engaged with the sport. The community loves football! The fantasy league has provided a whole new (and rewarding) dimension to its fandom. No longer is it tied down to rooting for a single team—instead, the whole league becomes fodder for its attention as it picks and chooses players from each of the 32 NFL teams. In addition, the fantasy football industry has coalesced around consistent formats for leagues, points, and key metrics (Figure 1-2). Terms like PPR, running back by committee, waiver wire, and flex are well understood, facilitating conversations among league owners. And with $1.18 billion bet in fantasy football leagues annually and a passionate fan base, fantasy owners have huge incentives to make informed decisions. When money or bragging rights are on the line, individuals invest time and energy into developing the skills and abilities to become data fluent.
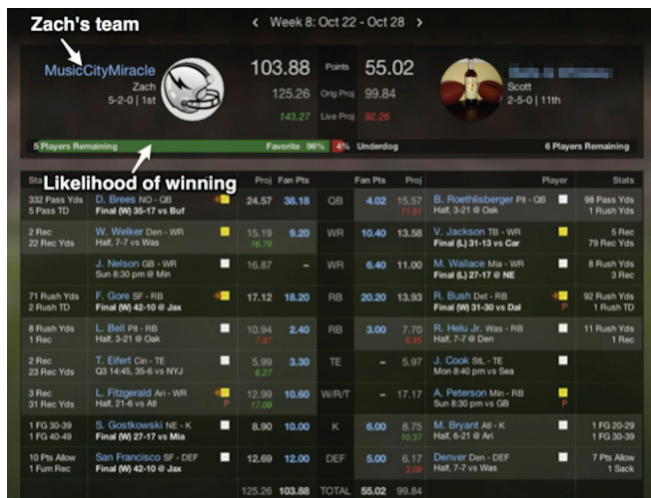

FIGURE 1-2 *Yahoo! Sports fantasy football game status page*

In short, these factors have brought data fluency to the masses. Millions of fans have learned how to read charts, grasp basic data concepts, and allow deeply embedded data to inform how they make decisions—all critical skills associated with quadrant one in our framework.

Producers of Data Products: U.S. News

The president of a prestigious nationally ranked university waited for the news. Ranked #20 in the previous year in *U.S. News & World Report's* Best College list, she felt enormous pressure from its large and influential alumni base to stay in the top 20. Though prepared to say that rankings didn't matter, she couldn't help but think about how she spent the past year touting its entry into the top 20 at every major fundraising and recruiting event. Not too far away from her office, high school seniors from the city's best school would use the rankings to guide their college selection.

The stakes are high. Economists James Monks and Ronald Ehrenberg found that a negative change in rankings would result in a decrease in applications, a decrease in applicant quality, and a drop in fundraising among select colleges and universities (Monks and Ehrenberg, 1999)[11]. A recent study further found that moving to the front page of the *U.S. News* ranking provides a significant increase in enrollment across all institution types (Bowman and Bastedo, 2009)[12].

Yet these rankings are controversial. Critics argue that the rankings show little variation across years—a measure that favors those who are consistently on top, institutions like Harvard, Princeton, Stanford, and a handful of other prestigious schools. Proponents, however, contend that the rankings help standardized college reporting and open visibility to the public with respect to financial data, teacher quality, student satisfaction and a host of other variables. Deciding where to apply and attend from approximately 2,400 public and private 4-year colleges is no easy decision.

Despite the controversy over the rankings, popular interest in the list has grown. For example, every September when U.S. News releases its Best Colleges list, the USNews.com website experiences massive volumes of visitors. In 2013, 2.6 million unique visitors generated almost 19 million views in a single day around the release of the rankings.[13] The online interest suggests people find the rankings to be an incredibly useful tool to sift through the vast amount of information available from colleges—information that is presented differently by each institution—to help guide college choice decisions.

So what can be learned from *U.S. News'* capability to bring data to a broad audience? How did it create such a successful formula? How did the *U.S. News* become the unofficial guide in the college attending process?

The *U.S. News* Best Colleges rankings list started as a reputational survey of college presidents. *People* Magazine had the "50 Most Beautiful People"; *Forbes* the "400 Wealthiest People." Both were top sellers. *U.S. News* and its leadership team understood the power of rankings. Bob Morse, director of Chief Data Strategist for *U.S. News*, and colleagues did not just recognize the power of lists to sell—they were also skilled data producers. They recognized that lists were a powerful tool that could consolidate vast amounts of information into a single, easily digestible metric.

In creating and disseminating the best college rankings, Morse and his team demonstrate four important principles in becoming a skilled data producer.

Categorize Information

Morse sorted higher education institutions into categories based on the Carnegie Classification of Institutions of Higher Education. Over time, new categories have been created by *U.S. News* to meet the demands of consumers. U.S. colleges and universities are classified into national universities, national liberal arts colleges, regional universities, regional colleges, best-value schools, up-and-coming schools, top public schools, and schools with the most inter-national students, for example. These categories provided college-goers with important information to begin their search.

Focus on Ranking and Comparison

Having established definable categories, other information could be used to rank and compare institutions within a given category. For instance, variables such as tuition and fees, total enrollment, acceptance rate, average freshman retention rate, 6-year graduation rate, class size, and average SAT scores could be used to sort and rank institutions. Ranking and comparison are easily acces-sible ways to present data.

Help Consumers to Understand What They're Seeing

Variables used to rank institutions were added over time. For example, the inaugural rankings in 1983 were based upon a simple survey to 1,308 college and university presidents asking them about the five best schools in their category. By 1988 the rankings included information on selectivity, faculty

strength, educational resources, and graduation rates.[14] Over time, Morse and his team introduced new variables to the models that produced the rankings. After the public understood the basics, *U.S. News* added complexity to refine the measures.

Create a Data Experience for a Diverse Audience

To aid students and families in the college choice process, *U.S. News* created My Fit Engine, an interactive tool that allows them to enter their preferences and find a list of matching schools. Although rankings appeal to students and families working through the college choice process, they have just as much (if not more) appeal to higher education institutions. In polling these institutions, *U.S. News* found that institutions are often concerned with historical trends and patterns. Accordingly, *U.S. News* worked with Juice Analytics to create Academic Insights, an online, interactive tool to analyze and explore data over time.

The great innovation of the *U.S. News* list is that it organizes vast amounts of information into a uniform format that is easily understood and then distributes that information via readily accessible print and online formats. *U.S. News* created a system where information on hundreds of institutions could be processed and reduced to a set of lists. Instead of sorting through countless websites and glossy brochures, families now had a quick way to reduce the amount of time invested in the college search process. In doing so, *U.S. News* demonstrates the skills central to quadrant 2 of our framework—the skills to create effective data presentations.

Organizational-Level Consumers: School District Woes

The meeting ended badly. For the last five years, the district had worked tirelessly to implement a new teacher evaluation system—and a corresponding database and reporting system to support its use. The district even allocated a large percentage of its $10 million grant from the federal government to hire "data coaches" to work with principals and teachers to develop familiarity with the data system and capitalize on its use. But the money ran out last year. The data coaches had to go, many to positions outside the district. The few that remained took central office jobs to work with the IT department on a wide variety of tasks, but with no unified purpose.

Now, one year later, the School Board called a special meeting to investigate the status and impact of this expensive endeavor. The central office team gave a dire report—unable to visit schools and without enough money to

provide ongoing professional development, it appeared that principals and teachers stopped using the system. The School Board members pressed the district about the functionality of the system by asking such questions as: *Is this a technological problem? Does your team not have the required staff to keep things running? Other than the data coaches, has anything else changed in the system since last year when most of the schools seemed to be using the data?* The district team responded that the data system not only functioned properly, but that it also had been improved to allow principals and teachers access to more reports and greater tools to help them explore and visualize the data.

"The system is way better now that we're working with the IT department," the leader of the department commented. "We're constantly updating what school personnel can do with it—the reports they can access, the tools they can use. We've made sure to send out weekly e-mails with all the changes and with reports that we believe are most useful. But we hear back from only a few who are really tech savvy and love data. We don't understand why the rest aren't using the system. It's clear that they don't like something about the system. They just never tell us what it is!"

As part of its presentation, the team provided information from a survey given to a random sample of schools in the district. The results were discouraging. Less than 10 percent mentioned using the system "regularly," and approximately 55 percent marked "never." Follow-up survey questions that explored the reasons for nonuse noted that a large majority of principals and teachers felt that the data systems and reports were "too complicated" to use. In addition, they felt that the district provided "too much information" and only 25 percent agreed that they provided "the right kind" of information.

After presenting the survey results, a group of teachers and principals that had been invited to the meeting were given time to discuss their experiences and answer questions. Their comments were revealing:

> One principal said: "I don't have time to open every e-mail I receive from the data office. And when I do, it seems like they're always changing the format. Sometimes it's a PDF, other times I have to open them in Excel."

> A teacher offered: "Half of the reports we get aren't even relevant to my school context and my student population."

> Another teacher asked: "How am I supposed to use teacher value-added information when I don't even understand how they calculate it?"

> A principal commented: "They're always changing how they do things. Some days I get on and find what I'm looking for, but then I go back a few days later

*and it's not there! And things that I get from their team hardly look like other
information I receive from HR or my area superintendent."*

*Another principal stated: "We're too busy to take the time to sit down as a
leadership team and look at the data."*

*A final principal said: "We have a hard enough time as 'insiders' making sense
of all the data . . . It's not like any parents are going to understand this stuff."*

After the teachers and principals concluded their remarks, the School Board
spent the rest of the meeting arguing over one major question: *Why did we
invest so much money in a system that isn't being used . . .and where do we go
from here?*

These challenges, which are explored in detail in Chapter 6, are all central to
organizational efforts to create a foundation for a common interpretation of
data presentations.

Organizational-Level Producers: Insurance Company Bottom Lines

Large companies spend millions of dollars a year on healthcare claims for
their employees. Health insurance companies are responsible for explaining
where all that money went, and why premiums go up annually, despite care-
fully designed programs to make a healthier workforce. Account executives
at insurance companies have a common challenge: how to tell this story to
their customers using data.

The goal is to show increased quality of health services being offered, bet-
ter health outcomes for the employees, and lower employer costs. Telling
a good story with actionable data would result in a contract renewal. Too
often, however, these data presentations are met with boredom and frustra-
tion rather than enthusiasm. A lack of a good story line is often the result of
disjointed data points pulled from a number of databases. Significant time
is spent collecting data and then trying to compare it to national, regional,
and industry benchmarks, and often little energy is put into telling the story.
Or conversely, much effort is put into telling a good story—with little data to
back it up. Either way, poor presentations are often followed up by requests
for different metrics and more data, which may or may not be helpful. And for
an account executive trying to close a deal, a poor presentation could mean
no contract renewal this month.

Consider the challenge of collecting wellness data on 250,000 employees
from a large company, and matching it with employee productivity, wellness,

and retention metrics at local and regional levels. First, data on the 250,000 employees is collected from many sources (employee, employer, and a variety of healthcare and wellness sources: hospitals, clinics, counseling centers, therapists, antitobacco programs, online programs, and even gyms). Some parts of the wellness program might be done in-house, whereas other campaigns like quitting tobacco and fighting depression were handled externally. Data collected on the 250,000 employees is then merged with that of 40 million other lives covered by this large insurance company—a large pool of data from which statistically significant comparisons could be made between those participating in a wellness program and those not. Data points came in daily—sometimes overnight via data sync between the employer's employee database and the insurance company's secure servers, and at other times via snail mail as in the case of gym reimbursement forms sent in by individual employees. Different departments collected different data points into 52 different databases. Extracting, analyzing, and presenting this data back to an employer in a useful and meaningful way is a challenge—demonstrating lowering healthcare costs, improving employee productivity and retention, and improving employee health and morale. Showing gains in these three areas would have to be contextualized for local, regional, and national trends.

Although this seems like an overwhelming task, large health insurance companies often have some of the most complete and comprehensive databases of employee claims and historical trends. Combined with top notch teams of statisticians and data scientists, it would seem pulling appropriate data for a large client presentation would be easy.

However, there are many breakdowns in the data ecosystem that prevent the systematic development and sharing of data presentations that can tell a compelling story. Perhaps these four challenges sound familiar:

- Data is stored in many separate, disjointed databases. Data is often collected from dozens of separate reports. Getting these reports may require a "ticket," queuing it up for a database administrator (DBA). Data retrieved might spark interest in another thread of reasoning and require another "ticket." And if reports are readily accessible without a DBA, significant time might still have to be spent logging into different data systems to collect the needed information.
- Consolidating all this data into a single PowerPoint presentation requires discretion. Synthesizing data and trying to make it look clean, professional, and visually engaging isn't a simple feat. Different reports

will have different metrics, and making sure the different metrics are communicated clearly in a visually accessible manner takes skill.

■ Telling a story is an act of creativity. Stringing these metrics into a story rather than a series of facts takes additional time and effort—and requires individuals who know both how to work technically and creatively.

■ Balancing a client's desire for detail on disparate topics with time constraints is a trick. Attending to a client's conflicting desire for detailed data with a 90-minute meeting constraint means that only the most meaningful data should be shared—but done so in an authentic manner—superficiality will not carry the day.

Telling a convincing story with actionable data sets the stage for a different type of meeting between organizations and their clients—a meeting where the client isn't even thinking about contract renewal or renegotiation but rather how to solve problems informed by thoughtful analysis. In the end, many companies struggle to tell a compelling story, grounded in carefully crafted data reports, to nondata-savvy consumers. These quadrant 4 skills, which are targeted in Chapter 7, are critical to developing an ecosystem for building and sharing data.

NOTES

1. Dragland, Åse, "Big Data—for better or worse." SINTEF, retrieved on July 22, 2013 from `http://www.sintef.no/home/Press-Room/Research-News/Big-Data--for-better-or-worse/`

2. Avanade, "Global Survey: Is Big Data Producing Big Returns?," published June 2012, retrieved on July 22, 2014 from `http://www.avanade.com/Documents/Research%20and%20Insights/avanade-big-data-executive-summary-2012.pdf`

3. Lucas Mearian, "By 2020, there will be 5,200 GB of data for every person on the Earth," *Computer World*, December 11, 2012, `http://www.computerworld.com/s/article/9234563/By_2020_there_will_be_5_200_GB_of_data_for_every_person_on_Earth?pageNumber=1`

4. Doug Laney, "3D data management: Controlling data volume, velocity and variety," META Group Research Note 6 (2001), retrieved on July 22, 2014 from `http://blogs.gartner.com/doug-laney/files/2012/01/ad949-3D-Data-Management-Controlling-Data-Volume-Velocity-and-Variety.pdf`

5. Office of Science and Technology Policy, "Obama Administration unveils 'Big Data' Initiative: Announces $200 Million in R&D Investments," retrieved on July 22, 2014 from `http://www.whitehouse.gov/sites/default/files/microsites/ostp/big_data_press_release_final_2.pdf`

6. Eric Sherman, "VC Bets $1.4 Billion on Big Data in 2012," *Inc.*, February 28, 2013, retrieved on July 22, 2014 from `http://www.inc.com/erik-sherman/vcs-bet-$14-billion-on-big-data-in-2012.html`

7. Tyler Bell, "Big Data: An Opportunity in Search of a Metaphor," *Radar*, February 10, 2011, `http://radar.oreilly.com/2011/02/big-data-metaphor.html`

8. CompTIA, "2nd Annual Big Data Insights and Opportunities" study, Retrieved on July 22, 2014 from `https://www.comptia.org/resources/2nd-annual-big-data-insights-and-opportunities`

9. Svetlana Sicular, "Big Data Is Falling into the Trough of Disillusionment," published January 22, 2013, retrieved on July 22, 2014 from `http://blogs.gartner.com/svetlana-sicular/big-data-is-falling-into-the-trough-of-disillusionment/`

10. Cindi Howson, "BI Adoption Remains Flat," *BI Scorecard,* April 1, 2014, retrieved on July 22, 2014 from `http://www.biscorecard.com/bi-adoption-flat/`

11. Monks, J. and Ehrenberg, R. G. (1999). The impact of U.S. News & World Report college rankings on admissions outcomes and pricing policies at selective private institutions (CHERI Working Paper #1). Retrieved June 1, 2014, from Cornell University, ILR School site: `http://digitalcommons.ilr.cornell.edu/cheri/1`

12. Bowman, Nicholas & Bastedo, Michael. "Getting on the Front Page: Organizational Reputation, Status Signals, and the Impact of U.S. News and World Report on Student Decisions." Research in Higher Education. August 2009, Volume 50, Issue 5, pp 415-436. Retrieved on June 1, 2014 from: `http://www-personal.umich.edu/~bastedo/papers/BowmanBastedo.ResHE2009.pdf`

13. U.S. News Breaks Online Traffic Record, published September 12, 2013. `http://www.usnews.com/info/blogs/press-room/2013/09/12/us-news-breaks-online-traffic-record`.

14 Luke Myers and Jonathan Robe, "College Rankings: History, Criticism, and Reform," *Center for College Affordability and Productivity*, March 2009, retrieved on July 22, 2014 from *Productivity* `http://www.centerforcollegeaffordability.org/uploads/College_Rankings_History.pdf`