HOW TO AVOID THE BIG BAD DATA TRAP

By Tamim Saleh, Erik Lenhard, Nic Gordon, and David Opolon

USING UNSTRUCTURED AND IMPERFECT big data can make perfect sense when companies are exploring opportunities, such as creating data-driven businesses, and trying to better understand customers, products, and markets. However, using poor-quality and badly managed data to make high-impact management decisions is courting disaster. Garbage in, garbage out, as the old technology mantra goes.

Data once resided only in a core system that was managed and protected by the IT department. In this small-data world, it was hard to get perfect data, but companies could come close. With big data, the quality of data has changed dramatically. Much of it arrives in the form of unstructured, "noisy" natural language, such as social media updates, or in numerous incompatible formats from sensors, smartphones, databases, and the Internet. With a small amount of effort, companies can often find a signal amid the noise. But at the same time, they can fall into the big bad data trap: thinking that the data is of better quality than it is.

The causes of bad data often include faulty processes, ad hoc data policies, poor discipline in capturing and storing data, and external data providers that are outside a company's control. Proper data governance must become a priority for the C suite, to be sure. But that alone won't get a company's data-quality house in order. Companies must adopt a systematic approach to what we call "total data-quality management."

The Impact of Big Bad Data

We regularly uncover contradictory, incorrect, or incomplete data when we work with companies on information-intensive projects. No matter the industry, often a company's data definitions are inconsistent or its data-field descriptions (or metadata) have been lost, reducing the usefulness of the data for business analysts and data scientists.

Using poor-quality data has a number of repercussions. Sometimes data discrepancies among various parts of a business cause executives to lose trust in the validity

and accuracy of the data. That can delay mission-critical decisions and business initiatives. Other times, staff members develop costly work-arounds to correct poor-quality data. A major bank hired 300 full-time employees who fixed financial records every day. This effort cost $50 million annually and lengthened the time needed to close the books. In the worst cases, customers may experience poor service, such as billing mistakes, or the business might suffer from supply chain bottlenecks or faulty products and shipments. The impact is magnified as bad data cascades through business processes and feeds poor decision making at the highest levels. (See Exhibit 1.)

In particular, companies routinely lose opportunities because they use poor-quality big data to make major executive decisions. From our experience with 35 companies, we estimate that using poor-quality big data sacrifices 25 percent of the full potential when making decisions in areas such as customer targeting, bad-debt reduction, cross-selling, and pricing. Our calculations show that revenues and earnings before interest, taxes, depreciation, and amortization could have been 10 percent higher if these companies had better-quality data.

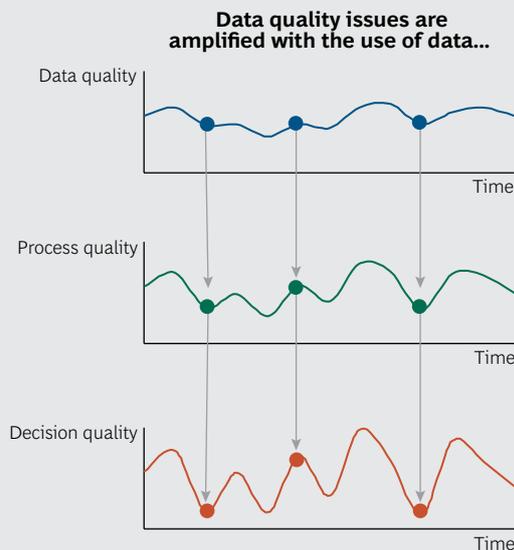A global financial institution conducted a big-data pilot project and identified what it thought was a pricing opportunity to increase margins by more than $50 million per year, or 10 percent of revenues. But the underlying data contained only invoice line items; it was missing important metadata about how the bank had calculated and applied fees. In the three months it took to correct the data quality issues and implement the pricing strategy, the company lost more than a quarter of its potential profits for the first year, equal to at least $15 million. It also lost agility in seizing an important opportunity.

In simpler times, companies such as this one could base decisions on a few data sets, which were relatively easy to check. Now, organizations build thousands of variables into their models; it can be much too complex to check the accuracy of every variable. Some models are so difficult to understand that executives feel they must blindly trust the logic and inputs.

How to Break Out of the Trap

As information becomes a core business asset with the potential to generate revenue from data-driven insights, companies must fundamentally change the way they approach data quality. (See "Seven Ways to Profit from Big Data as a Business," BCG article, March 2014.) As with other funda-

---

EXHIBIT 1 | Bad Data Destroys Value

Data quality issues are amplified with the use of data...

Data quality

Time

Process quality

Time

Decision quality

Time

...leading to major direct and indirect costs

Higher data-management budgets
- Storage costs
- Cleaning fees
- Manual work-arounds

Inefficient business processes
- Billing mistakes
- Supply chain bottlenecks
- Faulty products and shipments

Poor executive decision making
- Data disputes
- Lack of visibility
- Reduced agility

Source: BCG analysis.

mental shifts, mind-sets must be changed, not only technology.

However, companies frequently struggle to prioritize data quality issues or feel they must tackle all of their problems at once. Instead, we propose executives take the following seven steps toward total data-quality management, a time-tested approach that weighs specific new uses for data against their business benefits.

Identify the opportunities. To find new uses for data, start by asking, "What questions do we want to answer?" A systematically creative approach we call "thinking in new boxes" can help unlock new ideas by questioning everyday assumptions.[1] Other approaches, such as examining data sources, business KPIs, and "pain points," can be rich sources of inspiration. (See the "Big Data and Beyond" collection of articles for a sample of high-impact opportunities in a range of industries.)

After identifying a list of potential uses for data—such as determining which customers might buy additional products—prioritize the uses by weighing the benefits against the feasibility. Start with the opportunities that could have the biggest impact on the bottom line. Be sure to assess the benefits using multiple criteria, such as the value a data use can create, the new products or services that can be developed, or the regulatory issues that can be addressed. Then consider the uses for data in terms of technical, organizational, and data stewardship feasibility. Also look at how the data use fits in with the company's existing project portfolio.

Determine the necessary types and quality of data. Effectively seizing opportunities may require multiple types of data, such as internal and third-party data, as well as multiple formats, such as structured and unstructured. Before jumping into the analysis, however, the best data scientists and business managers measure the quality of the required data along a range of dimensions, including the following:

- *Validity,* the degree to which the data conforms to logical criteria

- *Completeness,* the degree to which the data required to make decisions, calculations, or inferences is available

- *Consistency,* the degree to which data is the same in its definition, business rules, format, and value at any point in time

- *Accuracy,* the degree to which data reflects reality

- *Timeliness,* the degree to which data reflects the latest available information

Each dimension should be weighted according to the business benefits it delivers, as explained in the previous step. We also recommend that companies use a multitier standard for quality. For example, financial applications require high-quality data; for bundling and cross-selling applications, however, good data can be good enough.
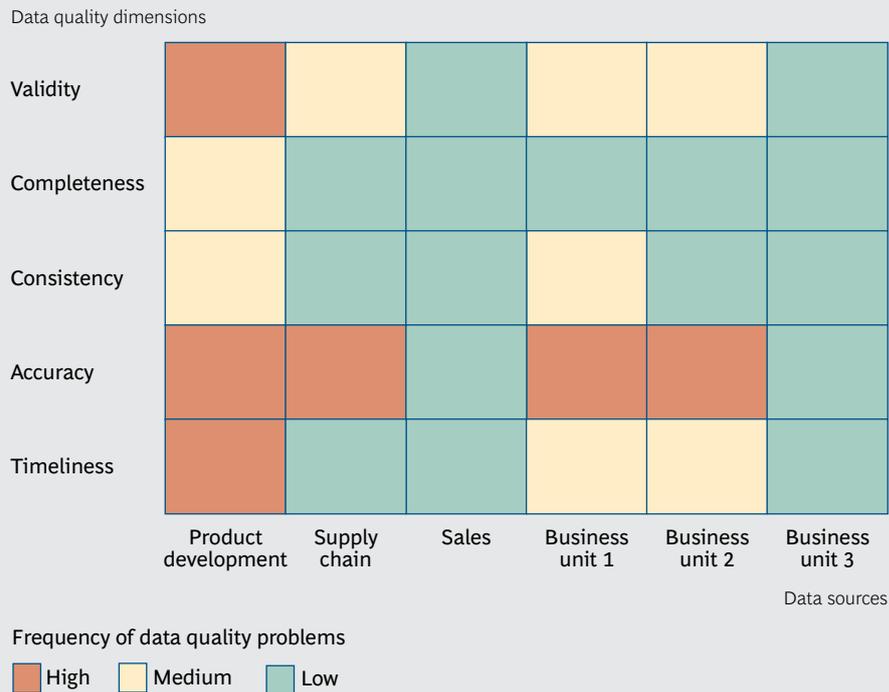
Define clear targets for improvement. The data-assessment process provides a baseline from which to improve the quality of data. For each data source, determine the target state per data quality dimension.

A gap analysis can reveal the difference between the baseline and target state for each data source and can inform an action plan to improve each data-quality dimension. Gaps can be made visible and tracked through a dashboard that color-codes performance for each of the major dimensions of data quality. (See Exhibit 2.)

Build the business case. Data quality comes at a price. To develop an argument for better-quality business data, companies must quantify the costs—direct and indirect—of using bad data as well as the potential of using good data.

Direct costs can include, for example, additional head-count expenditures that result from inefficient processes, cleanup fees, and third-party data bills. Indirect costs can result from bad decisions, a lack of trust in the data, missed opportunities, the loss of agility in project execution, and the failure to meet regulatory requirements, among other things.

EXHIBIT 2 | Dashboards Make Data Quality Gaps Visible

Data quality dimensions

Data quality dimensions	Product development	Supply chain	Sales	Business unit 1	Business unit 2	Business unit 3
Validity	High	Medium	Low	Medium	Medium	Low
Completeness	Medium	Low	Low	Low	Low	Low
Consistency	Medium	Low	Low	Medium	Low	Low
Accuracy	High	High	Low	High	High	Low
Timeliness	High	Low	Low	Medium	Medium	Low

Data sources

Frequency of data quality problems

■ High ■ Medium ■ Low

Source: BCG analysis.

The upside of high-quality data can be significant, as the prioritization of particular uses makes clear. For example, microtargeting allows companies to reach "segments of one," which can enable better pricing and more effective promotions, resulting in significantly improved margins. The more accurate the data, the closer an offer can come to hitting the target. For example, an Asian telecommunications operator began generating targeted offers through big-data modeling of its customers' propensity to buy. The approach has reduced churn among its highest-value customers by 80 percent in 18 months.

With costs and benefits in hand, management can begin to build the case for changing what matters to the business. Only then can companies put in place the right controls, people, and processes.

Root out the causes of bad data. Many people think that managing data quality is simply about eliminating bad data from internal and external sources. People, processes, and technology, however, also affect the quality of data. All three may enable bad data to accumulate. For example, we have seen companies spend enormous amounts of time and money cleaning up data during the day that is overwritten at night.

Certain types of data quality issues can and must be fixed at the source, including those associated with financial information and operating metrics. To do that, companies may need to solve fundamental organizational challenges, such as a lack of incentives to do things right the first time. For example, a call center agent may have incentives to enter customer information quickly but not necessarily accurately, resulting in costly billing errors. Neither management nor the data entry personnel feel what BCG senior partner Yves Morieux calls the "shadow of the future"—in this case, that entering inaccurate data negatively affects the overall customer experience.[2] Companies can also simplify data models and minimize manual interactions with data through approaches such as "zero-touch processing."

It may not be possible or economical to fix all data-quality issues, such as those associated with external data, at the source. In

such cases, companies could employ middleware that effectively translates "bad data" into "usable data." As an example, often the structured data in an accounts-payable system does not include sufficient detail to understand the exact commodity being purchased. Is an invoice coded "computing" for a desktop or a laptop? Workarounds include text analytics that read the invoice text, categorize the purchase, and turn the conversion into a rule or model. The approach can be good enough for the intended uses and much more cost effective than rebuilding an entire enterprise-software data structure.

Assign a business owner to data. Data must be owned to become high quality. Companies can't outsource this step. Someone on the business side needs to own the data, set the pace of change, and have the support of the C suite and the board of directors to resolve complex issues.

Many organizations think that if they define a new role, such as a data quality officer, their problems will be solved. A data quality officer is a good choice for measuring and monitoring the state of data quality, but that is not all that needs to be done, which is why many companies create the position of business data owner. The person in this role ensures that data is of high quality and is used strategically throughout the organization.

Among other responsibilities, the business data owner is accountable for the overall definition of end-to-end information models. Information models include the master data, the transaction data standards, and the metadata for unstructured content. The owner focuses on business deliverables and benefits, not on technology. The business ownership of data needs to be at a level high enough to help prioritize the issue of quality and generate buy-in but close enough to the details to effect meaningful change. The transformation needed is sometimes quite fundamental to the business.

Owners must also ensure that data quality remains transparent. Companies should have a target of 100 percent across all quality dimensions for customer data, such as names and addresses, and make that data accessible through a system such as a "virtual data mart" that is distinct from the storage of lower-quality data, such as reputation scores.

Scale what works. Data quality projects often run into problems when companies expand them across the business. Too many big-data projects cherry-pick the best quality data for pilot projects. When it's time to apply insights to areas with much higher levels of bad data, the projects flounder.

To avoid the "big program" syndrome, start small, measure the results, gain trust in effective solutions, and iterate quickly to improve on what works. But don't lose sight of the end game: generating measurable business impact with trusted high-quality data.

Consider the journey of an international consumer-goods company that wants to become a real-time enterprise that capitalizes on high-quality data. Before the transformation began, the company had a minimal level of data governance. Data was locked in competing IT systems and platforms scattered across the organization. The company had limited real-time performance-monitoring capabilities, relying mostly on static cockpits and dashboards. It had no company-wide advanced-analytics team.

To enable a total data-quality management strategy, the CEO created a central enterprise-information management (EIM) organization, an important element of the multiyear strategy to develop data the company could trust. The company is now beginning to centralize data into a single high-quality, on-demand source using a "one touch" master-data collection process. Part of the plan is also to improve speed and decision making with a real-time cockpit of trusted customer information that is accessible to thousands of managers using a standardized set of the top 25 KPIs. And the company is launching a pilot program in advanced analytics to act as an incubator for developing big-data capabilities in its business units and creating a path to

additional growth. Finally, it is creating a position for a business data owner who will be responsible for governing, designing, and improving the company's information model.

This multiyear transformation will be entirely self-funded from improved efficiencies, such as a projected 50 percent decrease in the number of employees who touch the master data and a 20 to 40 percent decline in the number of full-time information-management staff.

The Data Quality Imperative

Poor-quality data has always been damaging to business. But with the rise of big data, companies risk magnifying the impact of underlying inaccuracies and errors and falling into a big bad data trap.

Smart companies are beginning to take an end-to-end approach to data quality. The results from such transformations can be truly big.

NOTES
1. Luc de Brabandere and Alan Iny, *Thinking in New Boxes: A New Paradigm for Business Creativity* (New York: Random House, 2013).
2. Yves Morieux and Peter Tollman, *Six Simple Rules: How to Manage Complexity without Getting Complicated* (Boston: Harvard Business Review Press, 2014).

About the Authors

Tamim Saleh is a senior partner and managing director in the London office of The Boston Consulting Group and the global leader of the firm's big-data and advanced-analytics topic area. Tamim has extensive experience in areas such as business and IT transformation, big-data and technology strategy development, mergers and acquisitions, cost reduction, and outsourcing. You may contact him by e-mail at saleh.tamim@bcg.com.

Erik Lenhard is a principal in the firm's Munich office. He is a core member of BCG's Technology Advantage practice and a member of the firm's big-data leadership team. Prior to joining BCG, Erik was head of information management at Europe's largest cable operator. He is also highly experienced in data warehousing and agile software development. You may contact him by e-mail at lenhard.erik@bcg.com.

Nic Gordon is an associate director in BCG's London office. He has nearly 25 years of experience in financial services and has focused on developing data governance policies and building big-data capabilities. Nic has led global teams tasked with developing the IT strategy and data architecture for leading companies. You may contact him by e-mail at gordon.nic@bcg.com.

David Opolon is a principal in the firm's Boston office. He is a core member of BCG's Marketing, Sales & Pricing and Financial Institutions practices. David's expertise includes transforming pricing strategies and increasing companies' share of wallet using big data and advanced analytics. You may contact him by e-mail at opolon.david@bcg.com.

The Boston Consulting Group (BCG) is a global management consulting firm and the world's leading advisor on business strategy. We partner with clients from the private, public, and not-for-profit sectors in all regions to identify their highest-value opportunities, address their most critical challenges, and transform their enterprises. Our customized approach combines deep insight into the dynamics of companies and markets with close collaboration at all levels of the client organization. This ensures that our clients achieve sustainable competitive advantage, build more capable organizations, and secure lasting results. Founded in 1963, BCG is a private company with 82 offices in 46 countries. For more information, please visit bcg.com.