Appendix A
Measuring the Value of Data

An organization's data is a unique asset. It encapsulates knowledge about the organization and enables it to function and grow. It cannot be replicated by any other organization (Redman, 1996). Like other assets, data has value and it can be used to create additional value. Data also introduces risks. Its misuse can diminish the effectiveness and overall value of an organization. Despite the fact that data is routinely referred to as an asset, it is also recognizably not like other assets. Unlike money it is not consumed when used and it is not valued consistently. In many organizations, data is not valued in monetary terms at all. Many of the observations I have made about data quality measurement focus on the relationship between what data represents and how people understand it. If people don't understand the data they are using, they will not be able to maximize its value. And the value of data can change depending on other contextual factors. A Gartner analysis of data sprawl asked readers to engage in the following mental exercise about the ways in which the value of data can fluctuate:

> *Imagine that enterprise information is perfectly protected, and its accidental disclosure and loss is impossible. If a business user cannot access that information to make decisions, then the value of the information drops to zero. If the information is heavily replicated, with slight subtle variations in each replica, the reliability of the information, from a data quality and integrity perspective, is called into question, and thus the information's value drops again.*
>
> **(Glazer and Henry, 2012, p. 22)**

One of the recurring themes of data quality management is the question of how to determine the value of data. This appendix reviews what experts within information quality have said about the ways that data produces or contains value within organizations. I have not explored this theme in any depth in *Measuring Data Quality for Ongoing Improvement*. Instead I have started with the assumption that data is valuable and should be managed as an asset. The assertions documented below are the basis of that assumption. They provide food for thought as you build a business case for data quality measurement.

In *Data Quality for the Information Age*, Thomas Redman identifies six ways that poor-quality data can affect an organization's financial performance. They include: lowering customer satisfaction, introducing unnecessary costs, lowering job satisfaction and breeding organizational mistrust, impacting decision making, impeding reengineering (improvement) efforts, and impeding long-term business strategy (Redman, 1996, p. 6–11). He also points out that data persist in organizations—it fills the "white space in the organizational chart"—so a focus on producing and maintaining high-quality data can be "a unique source of competitive advantage" (Redman, 1996, pp. 12–13).

It is worth pointing out that Redman uses the term *unique* with its proper meaning: being the only one of its kind; unlike anything else. Any organization's data is unique to that organization—it is a reflection of the organization's history and knowledge. It cannot be purchased or replaced, and nothing can be substituted for it. In 1996, Redman pointed out that improving data quality is hard work, in part because "virtually everyone touches data" (Redman, 1996, pp. 12–13). In *Data Driven* (2008), he asserts that data is itself an active thing. Data and information are most valuable when they are "flying from place to place, helping complete a customer order here, contributing to a management report there, and stirring a new idea in someone's mind somewhere else" (Redman, 2008, pp. 29–30).

In other words, not only does everyone touch data, but data also touches everyone. And in doing so, it helps us understand its value.

Larry English devotes Chapter 1 of *Improving Data Warehouse and Business Information Quality* to describing "the high costs of low-quality data." He cites examples of direct and indirect costs incurred because of incorrect, incomplete or misleading information (English, 1999, pp. 7–12) and concludes: "the high costs of low quality data threaten the enterprise. …Information quality is a business necessity and information quality improvement is a business necessity" (English, 1999, p.13). Chapter 7 focuses on understanding these costs, which include process failures, scrap and rework, and missed opportunities (English, 1999, p. 209). He provides recommendations on calculating the costs of poor-quality information and of information value (English, 1999, pp. 221–225; 231–235). Importantly, English points out that part of the reason many organizations have poor-quality information is that they manage the systems life cycle, rather than the resource (information) life cycle (English, 1999, p. 207). This observation provides additional food for thought about the different assumptions of IT and the business when it comes to data quality.

In *Enterprise Knowledge Management: The Data Quality Approach*, David Loshin describes both the hard and soft costs of poor-quality data in operational, tactical, and strategic activities (Loshin, 2001, pp. 83–93). His categories form a framework that can be used to identify and assess the costs of low-quality data (and the corresponding benefits of high-quality data) within an organization. He defines soft impacts as those that "are clearly evident but still hard to measure"—things like difficulty in decision making and organizational conflict. In contrast, hard impacts—things like customer attrition, scrap and rework, and operational delays—can be estimated and measured (Loshin, 2001, p. 84). Hard operational impacts include the detection and correction of data errors, while soft operational impacts include public relations efforts ("spin"). Loshin also presents a process for using this framework to create an aggregated scorecard that "summarizes the overall cost associated with low data quality and can be used as a tool to find the best opportunities for improvement" (Loshin, 2001, p. 93).

"Assess Business Impact" is step four in Danette McGilvray's Ten Steps to Quality Data and Trusted Information.™ She presents eight techniques to assess the quantitative and qualitative business impact of data quality issues and includes templates to help with these analyses (McGilvray, 2008, pp. 163–198). The techniques include collecting anecdotes and examples of the impact poor data quality has had on the organization; creating an inventory of the current and future uses of data; drilling into data issues using the "Five Why's" to more fully understand the extent of business impact; creating a benefit versus cost matrix to understand the relative effects of data issues; ranking and prioritizing issues and remediation; using process diagrams to illustrate the effects of poor quality data on business processes; quantifying the direct and indirect costs; and using input from the various techniques, assembling a cost-benefit analysis that also includes a return on investment (McGilvray, 2008, p. 165). These techniques can be used alone or in conjunction with each other. The purpose of such an assessment is to build a business case and gain support for data quality improvement and to enable better decisions about where to invest in improvements. It has the additional benefit of improving organizational understanding of the data.

The DAMA Book of Knowledge takes a similar approach to describing data value both in terms of the benefits derived from the use of data and the costs of losing data or of having poor-quality data. DAMA recommends assessing potential changes in revenue, costs, and risk exposure (2009, p. 53). DAMA also points out that data can be valued in relation to what competitors might pay for data assets or in terms of the liabilities represented by information gaps.

Larry English provides an updated summary of the costs of poor-quality information in “Process and Business Failure: The High Costs of Low Quality Information,” Chapter 1 of *Information Quality Applied* (2009). These include well-known examples, such as the Mars Orbiter—$125 million plus lost scientific knowledge (English, 2009, pp.12, 19); quality concerns in the year 2000 election—4–6 million votes not counted, election decided by the Supreme Court (English, 2009, p. 20); and the audit failures at Enron causing billions of dollars in losses (English, 2009, pp. 8–9), as well as lesser known incidents like the Bureau of Indian Affairs spending $12.5 million on a software system and another $13 million on correcting records in the system (English, 2009, p. 14). English’s combined list of poor-quality software costs and poor-quality information costs includes references to more than 120 organizations and totals one and a quarter trillion dollars (English, 2009, p. 15). (Yes, trillion. That’s not a typo.) English concludes that, across numerous business sectors, these costs range from “20–35 percent of an organization’s operating revenue wasted in recovery from process failure and information scrap and rework” (English, 2009, p. 22).

It is clear from English’s numbers and the categories and techniques presented by Redman, Loshin, and McGilvray, that a necessary first step to improving data quality is to understand the value of data within an organization. Its value can be understood both negatively— through the costs of poor-quality data—and positively—through the benefits of high-quality data. Data’s quality has a direct impact on the value of data.