Executive Summary

Data quality is difficult to comprehend in its entirety, because of the diverse aspirations and actions collected under its broad umbrella. This includes standard technology and business practices that improve data, like name-and-address cleansing, record matching and merging, house-holding, de-duplication, standardization, and appending third-party data. Some of these tasks can be automated with software, while others—like entering data properly—are purely matters of business process.

Data quality is a complex concept that encompasses many technologies and business practices.

Given this complexity, it's no wonder misconceptions abound, like thinking data quality is a one-time action that results in perfection. To the contrary, data quality is a complex concept that encompasses many data-management techniques and business-quality practices, applied repeatedly over time as the state of quality evolves, to achieve levels of quality that vary per data type and seldom aspire to perfection.

Of the organizations TDWI surveyed, 82.5% continue to perceive their data as good or okay. However, half of the practitioners surveyed warn that data quality is worse than their organization realizes, which explains why the number of organizations with a data-quality plan doubled between 2001 and 2005. Many companies took action on data quality because compliance provided a swift kick in the pants. Other kicks came from initiatives for business intelligence, customer service, global supply chain, and IT system consolidations and migrations.

Most business people think their data quality is good, while half of technical people say it's worse than perceived.

Two-thirds of respondents have studied the problems of data quality, while less than half have studied its benefits. This indicates clearly that data quality initiatives are driven more by liability than leverage. In other words, organizations improve their data to avoid problems like direct-mail costs, misguided decisions, poor customer service, or faulty information in financial and regulatory reports. Of course, when these problems are fixed, data has greater leveragability. The benefits aren't completely overlooked, since most organizations surveyed claim a return on investments in data quality. Either way you look at it, the liabilities of poor-quality data and the leveragability of high-quality data should compel anyone to action.

Data quality is driven more by liability than leverage.

Data-quality products and practices are evolving quickly as they move from technical to business users, from point products to suites, from batch to real-time operation, from data profiling to quality monitoring, from US-centric to global, and so on. All these trends boil down to the fact that data quality is broadening beyond its departmental roots into enterprise-scope usage. While this broadening is good for the data, it's challenging for the organization, which must adjust its business processes and IT org chart to adapt to enterprise usage.

Data quality is broadening beyond its departmental roots into enterprise-scope usage.

Accomplishing anything with this kind of *enterprise data quality* (EDQ) requires close collaboration among IT and business professionals, who understand the data and its business purpose—collaboration made manifest in a data-governance committee or program. Data governance is rare today, but will proliferate as companies take data quality into broader enterprise use and move beyond mere stewardship. TDWI recommends data governance strongly, because it gives all data-management practices consistency, efficiency, and mandate as they reach for enterprise scale.

Data governance gives enterprise data quality the organizational structure it needs for collaboration between IT and the business.

Note that the most critical success factor for EDQ via data governance is *mandate*. Data stewards and governors must induce technical and business managers beyond their purview to change their processes and data when opportunities for data improvement arise. Without a strong mandate (supported by an attentive executive sponsor) to drive pragmatic changes, EDQ, data governance, and data stewardship deteriorate into an academic study of data.

Research Methodology

Report Scope. This report is designed for technical executives who wish to understand the state of data-quality initiatives today, as well as why they are evolving from departmental solutions to enterprise programs. The report describes a range of best practices for improving the quality of data, the drivers behind trends in data-quality practices, and how data governance is required for successful data-quality initiatives on an enterprise scale.

Terminology. TDWI defines *data quality* as the quality of data's content and structure (according to varying criteria), plus the standard technology and business practices that improve data, like name-and-address cleansing, matching, house-holding, de-duplication, standardization, and appending third-party data.

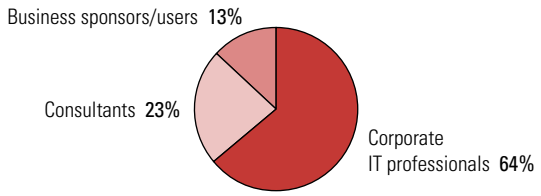
Survey Methodology. This report's findings are based on a survey run in late 2005 (with occasional references to a survey run in late 2001), as well as interviews with data-management practitioners, consultants, and software vendors. In 2005, TDWI sent an invitation via e-mail to the data-management professionals in its database, asking them to complete an Internet-based survey. The invitation also appeared on several Web sites and newsletters; 803 people completed all of the survey's questions. From these, we excluded the 53 respondents who identified themselves as academics or vendor employees, leaving the completed surveys of 750 respondents as the data sample for this report.

Survey Demographics. The majority of survey respondents (64%) are corporate IT professionals, whereas the remainder consists of consultants (23%) or business sponsors/users (13%). Hence, this market sample ably represents data-management practitioners and their business counterparts. Due to branching in the survey, some questions allow responses only from individuals who've had direct experience with data-quality or data-governance initiatives.

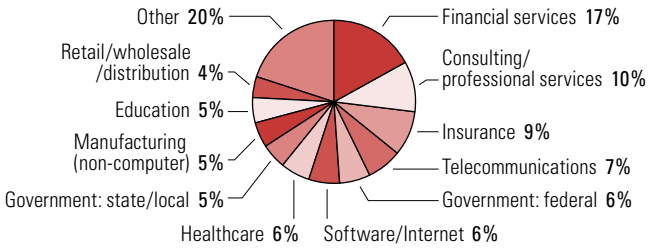
The financial services and insurance industries (26% combined) dominate the respondent population, followed by IT consultants (10%) and other industries (single-digit percentages). We asked consultants to fill out the survey with a recent client in mind. By far, most respondents reside in the U.S. (62%), trailed by Europe and Canada, respectively. Respondents are evenly distributed across all sizes of companies.

Demographics

Position

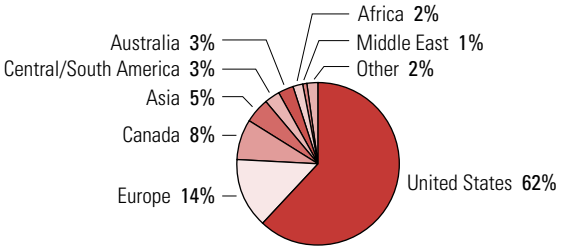


Industry

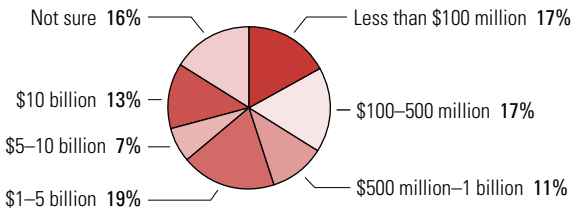


(The "other" category includes industries with less than 4% of respondents.)

Geography



Company Size by Revenue



Based on 750 qualified respondents.