

## Who's Responsible for Data Quality?

By  
*Thomas C. Redman, Ph.D.*

### Introduction

Last year in this space I wrote about the “social issues” that surround data quality and hinder people and organizations in their quality improvement efforts. In one article in particular (July, 2005) I explored the mistaken logic that “if it is in the computer, it must be IT’s responsibility.” According to this logic, responsibility for data quality should reside with IT.

To many organizations, the logic, even if only implicit, is impeccable. But assigning data quality to IT limits the organization’s range of actions and so dooms it to mediocre data. So where should responsibility for data quality reside? Should the “data quality function” be centralized or decentralized? And to whom do Data Quality Staff report?

### Two Basic Approaches to Managing Data Quality

As data quality practitioners are well aware, there are two basic approaches to managing data quality. The so-called first-generation approach aims to find and fix errors before they do much damage. This approach makes use of software tools to specify business rules, compare the data against those rules, and flag errant data. A first-generation approach can be led by IT.

The second-generation approach aims to prevent errors before they are made, at their sources. The second-generation approach is business-based. Though supported by software tools to help identify errors at their sources, measure quality, and establish control, second-generation work cannot be led by IT. The second-generation approach yields far higher quality data at far lower cost. And although the second-generation approach can be difficult to implement (because people and organizations must think and act differently), it is the only way to go.

### Sources of Data

As noted above, second-generation data quality management focuses on the sources of data. So where are the sources of data? Much data are created “in the business” – as manufacturing, customer service, development, financial, and other processes create data values in the course of day-in and day-out work. More data are created “in the supplier base,” as vendors, partners, and other third parties create data values that wind their way into the organization.

If errors are to be prevented, they must be prevented both in internal business processes and in the external supplier base. It therefore follows that organizations must identify their most important business processes and suppliers, name the responsible management teams, and invest these teams with accountability for the quality of data that their processes and suppliers create.

It is easy to glide over this last sentence. All too often I run into organizations that claim they actively manage their processes and suppliers. Yet, they cannot answer even my simplest questions:

- What are the ten most important processes/suppliers?
- Who is responsible for each?
- Who are the most important customers of each?
- What is the quality of data each is producing?

So let me be clear:

*The responsibility for data quality must reside in  
business process and supplier management.*

**Data Quality is an Inter-departmental and Business Responsibility**

Depending on the industry, many processes produce physical products or services as well as data. For example, the “Freight Delivery Process” may begin when a customer schedules a container pick-up, and ends when the item is delivered halfway around the world. Top-line metrics for this process may include on-time delivery, unit cost, customer satisfaction, as well as data quality. For large, cross-departmental processes such as Freight Delivery, it is particularly important that all (major) contributing departments sit on the process management team, because data errors most often crop up “in the white space” between departments.

We treat the “data modeling process” (i.e., the process of identifying entities and attributes, defining them, and creating the data structures needed to store them) just as we would any other business process. Data modeling is a perfect example of a process that crosses departmental boundaries. IT makes important contributions, but every department has its own *lingua franca*, which must be captured in the data. Therefore every department must contribute.

Many organizations already have a staff group responsible for the terms and conditions that govern their business relationships with suppliers. These groups often do not have detailed knowledge of how the supplied data are used to successfully lead the supplier quality effort. That understanding too, must reside within the business. For each important supplier, the Department most affected should head up the supplier team.

Data suppliers are not always external. For example, the team responsible for a data warehouse may manage the business processes that create data as if these processes were a data supplier.

**Decentralized, Specialized Efforts supported by Centralized DQ Staff**

To summarize, most of the day-in, day-out work of understanding data customer needs, measuring data quality, establishing statistical control, and completing data quality improvement projects is decentralized, in the hands of process and supplier management teams.

Supporting these teams is a small but potent Data Quality Staff. This group takes on tasks and issues that fall beyond the purview of process and supplier management.

Thus, for example, the DQ Staff:

- Lead the development and deployment of relevant Data Quality policies;
- Build the business case for the Data Quality Program;
- Ensure that all important data are covered by the process and supplier management structure;
- Define quality metrics and assemble the Data Quality Scorecard; and
- Develop deep expertise in the tools and techniques of data quality management, train others within their organization, and assist them as needed.

Now, where should this Staff group reside? There is no one answer that is suitable for every organization, but three pointers may help:

- First, the DQ Staff should report as high in the organization as possible;
- Second, they should report to the department where the “pain” due to poor data is greatest. That could be in Marketing, Operations, Risk Management, or Finance; and
- Third, they should report to the Executive who will lend his or her full weight to the effort.

As it stands now, many DQ Staff groups report to IT. This approach can work, and work well, if the CIO is an aggressive data quality champion and is viewed by the rest of the senior management team as a full partner. It will not work well when the IT department subordinates “data” under “technology,” or if the IT department is not a full partner of the rest of the business.

In time, organizations will recognize data as assets on par with other assets such as people and money. When that day comes, Chief Data Officers will be named, and will lead the work of the Data Quality Staff. I’ll discuss the roles and responsibilities of the Chief Data Officer next time.

Please send in your thoughts and comments to [tomredman@dataqualitysolutions.com](mailto:tomredman@dataqualitysolutions.com)

### About the Author



**Dr. Thomas C. Redman**, Co-Founder of the IAIDQ, is President of Navesink Consulting Group, based in Little Silver, NJ.

Tom holds a Ph.D. in statistics from Florida State University and is an internationally renowned lecturer and the author of numerous papers, and has written three books, *Data Quality: The Field Guide*, (Butterworth-Heinemann, 2001), *Data Quality for the Information Age* (Artech, 1996) and *Data Quality: Management and Technology* (Bantam, 1992). Tom holds two patents.

© 2006 Navesink Consulting Group