Putting Quality Process in Place to Exploit Technology

Prepared by:
Larry English, Information Impact International & the DataFlux Corporation

Putting Quality Process in Place to Exploit Technology

Figures

Executive Summary

Data or information quality is one of the latest “hot topics” in business today. The concept of faulty or nonquality data – once a problem that resided primarily with IT staff – is moving from the server room to the boardroom. The effects of bad data can at best be a hindrance to corporate efforts such as marketing, sales, business intelligence, customer support and finance. At worst, inconsistent, inaccurate and unreliable data can have a calamitous effect on data-driven efforts within the enterprise.

Part of the increased interest in data and information quality in business today is because of the growing role of data quality technology used to address faulty or nonquality data. Through this software, businesses can indeed build better data sources, which can then drive enterprise applications such as customer relationship management (CRM) or enterprise resource planning (ERP) systems. The promise of applications such as CRM and ERP is to provide access to information from throughout the enterprise. Without good data, that promise will go unfulfilled.

However, software alone will not – and cannot – solve the quality problem within a company’s information resources. The problem requires a solution that spans the roles, responsibilities, policies and procedures that dictate the acquisition, maintenance, dissemination, and disposition of data. Without addressing each of these facets, especially the processes behind the data, any data quality initiative is in jeopardy from the beginning.

This paper will explore the process improvement technologies that are commonly used to address information quality. Originally developed to manage the quality of manufactured products, these techniques have similar applicability in the field of data management as they provide critical insight into the management techniques, organization strategies and process improvements necessary to make quality a corporate imperative. Companies can then apply these techniques in tandem with data quality technology to develop a true enterprise solution for information quality.

Root Cause(s) of Nonquality (or Poor Quality) Data

Over the past few decades, companies have increasingly realized that creating – and maintaining – quality data is a tricky business. Customers change addresses without notifying vendors. Product numbers shift as product lines morph. Companies merge business applications from a host of sources – and often import existing quality problems into the new target database. Some industry estimates show that data quality problems cost U.S. businesses more than $600 billion per year.[1] Other studies show that the cost to the U.S. economy is closer to $1.5 trillion or more per year.[2]

But where do the problems start? Where does this massive amount of incomplete, faulty or otherwise unusable data originate? The answer often lies within the processes or procedures intended to guide how an organization collects and maintains its data. To create an effective data management program – one that allows you to collect and maintain high-quality data – you need to address the business processes and the data management processes.

Data problems: Real-life Headaches. Costly Consequences.

> **"Oh, I hate it when it does that. [Computer shows that a required address field was left blank.] I'll just enter '1-2-3' for the address and keep going to the next screen."**
>
> **Unnamed auto parts store sales associate**

This quote occurred during a transaction at a nationwide auto parts store. A customer had purchased a new battery, and the sales associate was entering information about the sale into the point of sale computer. After getting the customer's phone number and attempting to go to the next screen, the sales associate received a system prompt to ask the customer for an address. At that point, the clerk – frustrated that he could not complete the sale and help other customers – just entered "123" into the address field, clicked "OK" and went to the final screen.

At that point, the customer database of this retailer received a new, partially-valid customer record. The name and the phone number of the customer were valid, but the address was an insignificant fragment. If this same dynamic happened at other stores nationwide, the company could receive thousands (if not millions) of invalid or incomplete customer records each year.

Addressing data quality problems requires a mixture of technology and process improvement to realize sustainable quality. In this example, the retailer needs to determine:

- Is address collection at the point of sale the optimal time to receive this data?
- If so, do clerks and sales associate understand the procedure and protocol for data entry at the point of sale?
- Can the company integrate data quality technology within the existing systems to identify nonquality data in real time and automatically correct it before it enters the first database?

By answering these questions, this retailer can solve what may be a prevalent problem within the organization. The ultimate solution will most likely require a combination of a process improvement methodology and data quality technology.

---

[1] "Data Quality and the Bottom Line: Achieving Business Success through a Commitment to High Quality Data." The Data Warehousing Institute. Report Series 2002.
[2] L. English, "Reversing the High Cost of Low Quality Information", *Proceedings of The 13th Information Quality Conference*, San Diego, CA, 2003, p. 18.

Poor Quality Information = Costly Consequences

The best case scenario in organizations with unmanaged information quality is that they are wasting from 10% to more than 20% of their operating revenue (or operating budget) in recovering from process failure and in "information scrap and rework"[3] caused by defective information. Information scrap and rework includes all costs associated with hunting and chasing information, redundant data entry, verifying and correcting data, compensating alienated customers, handling returns, wasted materials and fines for noncompliance. Indirect costs come in the form of missed opportunity and in lost customer lifetime value.

Left unchecked, information quality problems can lead to enterprise failure. More than a few companies have gone out of business in part because of poor quality information.

Quality Management Methodologies and Information Quality

In the first century BC, Publilius Syrus wrote, "From the errors of others, a wise man corrects his own."

This is true for information quality management today. The principles, processes and methods for information quality management already exist and have existed for decades. The same quality management systems that W. Edwards Deming (14 Points of Quality), or Joseph Juran (Quality Planning, Improvement and Control), Philip Crosby (14 Steps of Zero Defect Quality) or Masaaki Imai (Kaizen®) have applied to manufacturing also apply directly to information quality. The reason? Data is a "product" of business and manufacturing processes just as much as a Toyota Corolla, VIN 12…69, is a product of the Toyota Motors manufacturing processes.

What is the lesson here? Sound quality management methodologies already exist. We must learn from them and apply them to information.

Six Sigma is a re-packaging of the same quality principles and techniques and has gained popularity and support because it addresses quality as a business economic problem – and when rigorously applied produces dramatic financial results. Organizations applying Kaizen, Deming's 14 Points and other sound quality systems also yield dramatic results when applied rigorously.

These valid quality management systems led the world in the first quality economic revolution, eliminating waste in manufacturing and transforming the rules of competition in the Industrial Age. This represented the maturing of the Industrial Age, from accepting "scrap and rework" as a normal part of doing business, to "design quality in" to eliminate defects and the costs of scrap and rework.

Four Fundamental Principles of Valid Quality Management Systems

Common to each valid quality management method – those that produce business value by eliminating the waste of nonquality – are four fundamental principles. TIQM is a comprehensive quality management system that applies Deming's 14 Points of Quality, Kaizen and the Baldrige Criteria to information. The four principles that guide this effort are: Customer Focus, Process Improvement, Scientific Methods and Management Accountability.

---

[3] Larry English, *Improving Data Warehouse and Business Information Quality (IDQ&BIQ)*, NY: John Wiley & Sons, 1999, p. 210.

Customer Focus

Every valid quality system starts and ends with the customers. It begins with understanding the customers' requirements for the product or service, establishing quality standards to meet them and ends with measuring and confirming customer satisfaction

The technique quality systems use is getting suppliers and customers together to understand those needs and to work as a team to "consistently meet customer expectations." Customers may be external or internal. On the assembly line, the next process is "my" customer. Deming aptly puts it when he says, "the consumer is the most important part of the production line."[4]

In information quality, information producers and their managers must know their information customers, who are knowledge workers who depend on their data. They must understand their information quality requirements, such as accuracy, timeliness, non-duplication and consistency of data in redundant databases, and the costs caused downstream by defective data.

Because the information customer is the most important part of the "information value chain," information professionals need a new word to refer to them. The word "user" does not adequately convey the importance of those who require information to perform the value work of the enterprise. Peter Drucker, the renowned management guru coined the term "knowledge worker" as early as 1960 as the evolution of the factory "worker" when he foresaw the impact of information technology on business and society. In "The Information Quality Act" that requires US federal agencies that disseminate influential information to the public, to provide quality information, refer to the recipients as "information consumers."

Many organizations that are serious about applying quality management to information are referring to business personnel now as "knowledge workers," "information consumers," "information producers," "business professionals" (parallel with "systems professionals") and "business partners."

Process Improvement

Every valid quality system concentrates on "process improvement" as the core process of quality management. The earliest quality systems in manufacturing implemented "inspection" and "corrective maintenance" or "scrap and rework" activities to take the *normal and accepted* degree of nonquality products produced by the manufacturing processes and fix them.

Assessment processes are cost-adding; their value is in confirming that processes that produce a product are either in control or that they require improvement. The value basis comes when the processes are improved and controlled to eliminate defects and the subsequent costs of scrap and rework.

That process improvement technique used by all valid quality systems is The Shewhart Cycle of PDCA (Plan-Do-Check-Act) and its variation in Six Sigma, DMAIC (Define-Measure-Analyze-Improve-Control).

Scientific Methods

Again, all valid quality systems use demonstrated, proven scientific methods, which in addition to the PDCA or DMAIC, include, but are not limited to:

- Cause and Effect Analysis and Ishikawa Charts (Fishbone Diagrams)
- The SIPOC (Supplier-Input-Process-Output-Customer, a tool to understand customer requirements and customer-supplier relationships)

---

[4] W. Edwards Deming, *Out of the Crisis*, Cambridge: Massachusetts Institute of Technology Center for Advanced Engineering Study, 1986, p. 26.

- QFD (Quality Function Deployment) and "The Voice of the Customer" that involve the customer in the design of products and services, not just in requirements definition
- Benchmarking, both internally and externally to compare process capabilities
- SQC (Statistical Quality Control) and Quality Control Charts, that measure quality over time for the purpose of assuring processes are "in control," producing quality products consistently

Information quality management can apply these very same tools.

Management Accountability

Accountability must be applied at the right level to produce quality, whether in manufacturing or in information production processes. The accountability for information quality resides with the managers of the processes that create and update information. This accountability must be a part of every manager's job description and part of managers' performance measures. Managers must be accountable for the quality of information so all downstream information consumers can perform their processes right the first time as a result of having complete, accurate, non-duplicate information on a timely basis.

Information Quality Management Systems

When organizations apply these four principles to data and information, they will create a second economic revolution. These valid quality management techniques pave the way forward to eliminate waste in information processes, transforming the rules of competition in the realized Information Age. This results in the maturing of the Information Age, from accepting "information scrap and rework" as a normal part of doing business to "design quality in," eliminating defects and the costs of "information scrap and rework."

Total Information Quality Management (TIQM®)

TIQM, developed by Larry English, uses the principles and techniques of quality management gurus, such as Deming, Juran, Crosby, Masaaki Imai (Kaizen), who led the world in the first quality revolution. TIQM applies the four quality management principles mentioned above, and applies them to address the unique issues in quality of "intangible" information versus tangible products.

TIQM consists of six processes for information quality management, illustrated in Figure 1. The umbrella process P6 "Implement an Information Quality Environment" provides for the cultural transformation required to sustain an IQ function by making it a value system and habit within the culture.

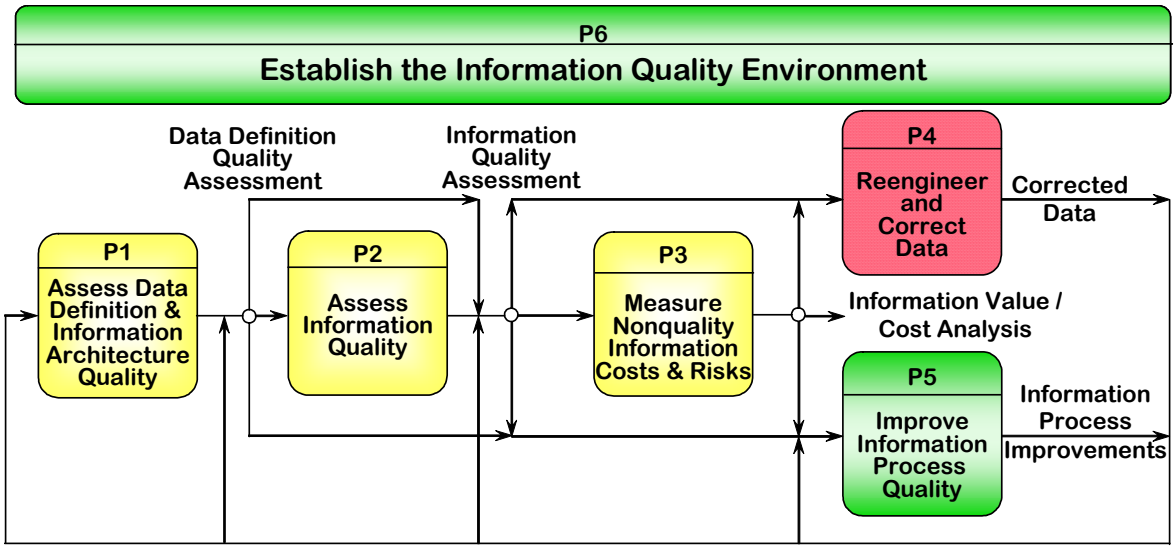

Figure 1: Six Processes of Total Information Quality Management[5]

The other five processes are part of the activities of quality assessment (P1 and P2), cost measurement (P3), corrective maintenance and data movement process control (P4) and process improvement (PDCA) (P5).

- P1 assesses the quality of "information product specification data," which includes data names and abbreviations, data definition, valid value sets and business rules, along with the quality of data models (what the enterprise needs to know) and the physical database designs (the knowledge repositories). If data definition quality is deficient, you must conduct a P4 to correct the data definition to create a common language and understanding of the data *and* a P5 to improve the data definition process.
- P2 assesses the quality of the data and information in the various quality characteristics important to knowledge workers.
- P3 measures the most important metric to executive management – the costs of the waste caused by defective information. Use this to prioritize P4 and P5 initiatives.
- P4 describes how to conduct data cleansing activities efficiently as a one-time activity, coupled with a P5 process improvement initiative. It also describes how to put audits and controls in processes that move and transform information where data movement is required.
- P5, the core competency process, conducts the Plan-Do-Check-Act cycle to deliver business value by eliminating waste and preventing defects. The steps are:
  - Define an improvement project where pain and high costs of nonquality information are being felt
  - *Plan* an improvement by first analyzing root cause(s) of the information quality defects and then defining improvements whose goal is to prevent recurrence of the defects
  - *Do* implement the improvements in a controlled environment so you can assure they solve the problem and prevent recurrence
  - *Act* to put the process in control to "hold the gains" and roll the improved process out to the rest of the organization

[5] Larry P. English, "Improving Information Quality: Processes and Best Practices for Business Performance Excellence" seminar, (from English *IDW&BIQ*, p. 70). Brentwood, TN: Information Impact International, Inc. 1994-2004, p. 34.

- P6 begins and continues the journey of transforming the culture to a value system of valuing ones information customers, a mind-set of excellence in all products including the information product one produces, and a habit of continuous improvement of all processes, beginning with the most important.

Six Sigma Quality System

Six Sigma is a Quality Management system created by Motorola to eliminate costs during a time of intense competition. Jack Welsh, former CEO of General Electric, saw its value and implemented it vigorously at GE. Six Sigma is both a cultural transformation and a rigorous process improvement method, recasting the PDCA into DMAIC, embedding measurement into projects that focus on improvement. Figure 2 illustrates the DMAIC Improvement Model and the DMAIC Phases.

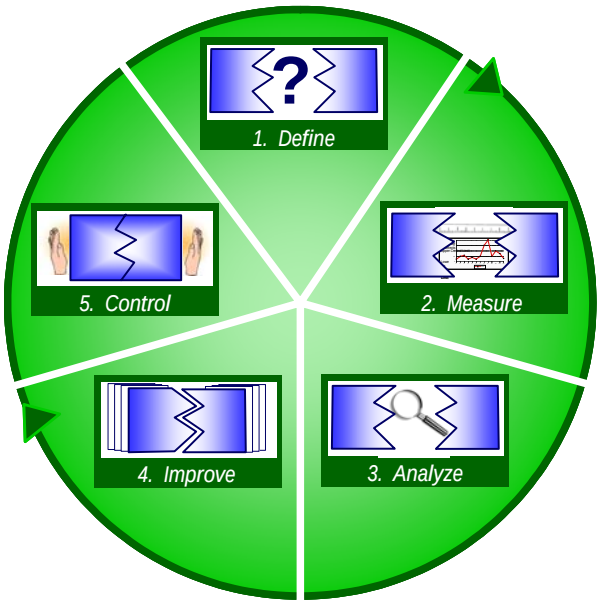

Figure 2: Six Sigma DMAIC Improvement Model[6]

Six Sigma has become popular because it focuses strongly on a balance of:

- A focus on the customer. The first three of the original six steps were: 1) Identify your product or service, 2) Identify the customer(s) for your product or service; determine what they consider important, and 3) Identify your needs to provide the product/service so that it satisfies the customer.[7]
- A value proposition of decreasing the costs of nonquality.
- A focus on process improvement and process design or redesign as the means of addressing nonquality. DMAIC does discuss scrap and rework as a means of "improvement." It is centered around improving, controlling and performing processes to approach six sigma-level quality based on what customers care about.

---

[6] Peter Pande, Neuman, and Cavanagh, The SIX SIGMA Way, NY: McGraw-Hill, 2000, p. 39.
[7] "Utilizing the Six Steps to Six Sigma," Motorola University, 1992, as cited in C. Sengstock, Jr., *Quality in the Communications Process*, Motorola University Press, 1997, p. 11.

- Active involvement of top management who understand and champion the imperatives of quality principles and processes to accomplish the value proposition. When Jack Welch saw the value of Six Sigma, he quickly became the champion, urging "his top lieutenants to become 'passionate lunatics' about Six Sigma."
- A rigorous set of processes and techniques to measure, improve and control the quality of the product, service or information based on what is important to the customer. The processes of measurement and techniques for improvement are not new to Six Sigma. According to one of the Six Sigma experts at Motorola, those techniques were not invented by Motorola. Indeed, they are basically the same best practices that were developed by quality pioneers like Shewhart, (PDCA process improvement cycle), Ishikawa ("Fishbone" Cause-and-Effect" diagrams), Statistical Quality Control Charts, on down the line… It is the rigor of applying them for the purpose of improvement and customer satisfaction that needs to be achieved.
- Six Sigma improvement projects are sponsored, led and coached by personnel who are certified in the Six Sigma measurement and improvement techniques, Master Black Belts, Black Belts or Green Belts, depending on the size and complexity of the projects."[8]

Six Sigma and TIQM

TIQM and Six Sigma are compatible. In fact, several organizations perform TIQM processes in Six Sigma DMAIC projects. Figure 3 shows the relationship of the overall Six Sigma Roadmap for implementing quality into the culture of the enterprise and how the six processes of TIQM correlate to them.[9]

Six Sigma Roadmap	TIQM Processes	
1. Identify core processes and key customers	P2S3	Identify information value and cost chain
	P1S3	Identify information stakeholders
2. Define customer requirements	P2S2	Plan IQ objectives, measures and tests (based on customer expectations)
3. Measure current performance	P1	Assess data definition and architecture quality
	P2	Assess information quality
4. Prioritize, analyze and implement improvements	P3	Measure nonquality information costs and risks
	P5	Improve information process quality
5. Expand and integrate the Six Sigma system	P6	Establish the Information Quality environment

Figure 3: Mapping TIQM Processes to Six Sigma Roadmap[10]

[8] Larry English, "Six Sigma and Total Information Quality Management (TIQM)," *DM Review*, October, 2004.
[9] For a detailed map of TIQM Processes and Process Steps map to Six Sigma's DMAIC Phases and Steps, please see "Six Sigma and Total Information Quality Management (TIQM)" in the October 2004 issue of DM Review magazine.
[10] *Ibid*.

Conclusion: Successful Organizations Exploit Information Quality Technology on the Framework of Sound Quality Management Processes

Organizations that acquire information technology without having a sound information quality management system may get some benefits, but they will not fully exploit the value of the technology. No technology alone can cleanse or correct all data defects. Some defective or missing data about events, once captured, may not be able to be corrected because there is no way to get back to the actual event.

But organizations that implement effective information quality methods based on sound quality management principles will exploit the full value of their information quality technologies.

The Processes Behind Data

Regardless of its current state, the origin of most corporate data is a human source, such as an employee, a partner organization or a customer. When people create defects, they are following a defective process. If people do *not* follow the process, then there is a broken management process of training or performance measures.

Because data is such a critical component of any successful business, people typically don't knowingly or willingly create bad or unusable data. Nonquality data could be the result of rushed data entry, insufficient training, ambiguous or non-existent data definition, or other environmental conditions, such as quotas or performance measures.

Companies worldwide have implemented process improvement methodologies designed to address these questions. What processes are not working optimally? And what can the company do to fix them? Although originally intended to decrease defects in the manufacturing process, these methodologies can apply to corporate information. From Six Sigma to Kaizen to TIQM, the process improvement methodologies share a common theme. There are three primary steps to any improvement process:

- Identify the broken processes and analyze root cause(s)
- Improve the process
- Control the process

Based on these fundamental process steps, you can implement a comprehensive enterprise data management process for dealing with complex data issues. After improving the process, you can then control the level of information quality over time. The result is the Analyze-Improve-Control cycle.

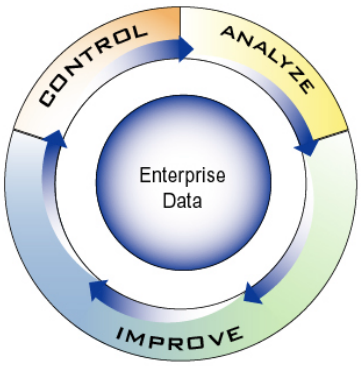

Figure 4: The Process Flow for Data Management.

As any company who has implemented quality improvement methodologies knows, you must find and eliminate the root cause(s) of the problem – not just "rework" (correct) the defective product. This section explores the Analyze-Improve-Control cycle, and how you can augment a process improvement program with data management technology.

Analyze: Study the Process Effect, Analyze Root Cause(s) – and Plan Improvement(s)

The first step in this data management methodology involves discovering the root cause(s) of the problem and defining a path to improvement. At this point, you can examine the data to find trends or anomalies that point to faulty or improper data collection or maintenance techniques. In Six Sigma, this is typically defined as "defects per million opportunities," or DPMO. Other process improvement methodologies use percent of defects or measure the costs of the defects.

To assist your organization throughout the analysis process, you can utilize data profiling technology to ascertain the strengths and weaknesses of your data. Profiling – also called discovery or inspection – can identify quality problems that may exist. This initial analysis measures all aspects of data quality, including completeness, accuracy, consistency and duplication of your data. A data profiling effort can also examine whether the data deviates from enterprise-wide data standards.

For instance, suppose you have a customer database that includes a column for phone numbers. All new orders entered by sales associates should include a valid phone number. However, a profiling report shows that 35% of customer records contain an incomplete, null or invalid phone number. From this analysis, you know that the process for entering some contact information (particularly phone numbers) needs to be improved.

In this example, you can further examine the data to find trends that may help you find the root cause of the failure. Are there common themes in the incorrect data that show that the sales associates need a quick refresher course in data entry procedures? Could sales representatives benefit from a real-time data quality technology that cross-references names and addresses to augment the record with phone numbers automatically? By answering questions such as these, you can design a program to improve the processes behind the data.

To truly analyze and discover the real root causes, you must bring together subject matter experts (SMEs) who perform the processes of creating and updating the data. Also involved will be SMEs who understand the applications that automate the processes and those who are responsible for facilitating the data definition and who create database designs. To analyze root cause requires the knowledge of those who perform the pertinent processes and who understand the problems in the processes. When empowered, they will help identify the root causes and improvements.

A trained and skilled process improvement facilitator is required to lead the team in the discovery of root causes. This person must know how to differentiate between precipitating causes and root causes. The improvement facilitator must help create a non-blame, non-judgmental atmosphere to protect people from the fear of being blamed for the defective data. If you have defective data, it is caused by defective processes—*not* defective people.[11] People must *never* be blamed for defective data or the broken processes that produce it.

[11] Larry English, *IDW&BIQ*, p. 288.

Improve: Fix the Process – and Eliminate Defects

Once the specific data quality issues and their root causes have been identified in the analyze step, you can then define and test process improvements to eliminate the cause(s) of defective data. If information consumers do not trust or use the existing data, you may perform corrective maintenance on the data to eliminate as many of the defects as feasible to enable the knowledge workers to trust it.

Error-proofing techniques to improve the process may be readily apparent given an understanding of the cause(s). One common method is to assign clear procedures to improve data entry processes to improve the quality of corporate information. Or if the cause is lack of knowledge, you can conduct intensive training and institute performance measures to help improve the data collection process. If, however, the cause is performance measures that provide incentives for speed of order capture or require quotas, you must address a balance of quality measures based on the cost of defects to downstream processes and to the end customer.

Data management technology can assist in the process improvement procedure by providing an automated way to verify data as it is created and as it moves through the system. You can utilize this technology to:

- Validate information during original data capture – Real-time data quality technology that provides “defect-prevention” capability can provide near-instantaneous feedback that data meets expected standards or protocols.
- Perform data checks as information flows from application to application – By utilizing data quality workflows that provide control of data movement, you can establish process checkpoints to guide management of data movement.

Control: Manage the Process – and Set Goals for Quality Data

After improving the processes that create and update your data, design quality controls into the process to ensure that you maintain high levels of data quality on an ongoing basis. This helps you assure zero-defect and/or high-priority data elements within the organization that are mission-critical to the success of the organization. Then, assess the quality of zero-defect data regularly, to provide feedback to the source personnel and to assure the data consistently meets its information consumers’ needs.

To effectively monitor and control your data, you can enact a few mechanisms that routinely check the health of your corporate information, including:

- Ongoing reporting and analysis of potential problem areas
- Alerting mechanism that recognizes out-of-control data records – and automatically flags the process owner or responsible party
- Trending of data to recognize cyclical variations to allow business analysts to view a historical view of data problems

With a data monitoring program, you can create goals for the company moving forward to maintain high levels of data quality. By comparing actual results to your goals, you can then adjust and refine processes to improve overall data integrity.

Getting Started

A pioneer in data management technology and services since 1997, DataFlux provides solutions designed to improve the consistency, accuracy and reliability of an organization's business-critical data. DataFlux frequently works with clients to implement data management solutions within a framework of overall process improvement throughout the enterprise.

By integrating process improvement methodologies into its core technology offering, DataFlux allows companies to elevate the role of data management into an ongoing corporate activity that can develop and maintain high levels of data quality. DataFlux has designed and implemented a three-phase data management process – Analyze-Improve-Control – that helps companies build and retain high-quality data sources.

The Growing Role of Data Management

The process of data management begins with a discovery or data profiling phase that asks one critical question: What points of data collection might have relevant, useful information for your data-based applications and initiatives? Once you begin to understand your data, you can correct errors and improve the data, allowing you to exploit quality information to build more effective CRM, ERP, data warehousing and other applications. And after establishing rules to build data, the final phase allows you to control data quality over time, creating an environment where any decline in data integrity can be addressed and corrected before bad data can impact your operations.

DataFlux technology provides the tools for this approach through five building blocks:

- **Data Profiling** – Discover and analyze data discrepancies
- **Data Quality** – Reconcile and correct data and improve the processes that create it
- **Data Integration** – Integrate and link data across disparate sources
- **Data Augmentation** – Enhance information using internal or external data sources
- **Data Monitoring** – Check and control data integrity over time

With this technology, DataFlux can help you improve every facet of your data management program – from initial data discovery and quality assessment to process improvement and process control with ongoing monitoring of data quality.