

The Essentials of Information Quality Management

By Larry English

Information quality (IQ) and data quality (DQ) are becoming the new buzzwords. Many organizations have IQ or DQ projects under way in some shape or form, whether in the context of data warehousing, CRM or e-business. Software providers are quickly developing new data quality products. Consultants are jumping to hang IQ or DQ on their list of services.

However, I must raise a warning flag! There are many attempts to apply the words “data quality” or “information quality” to practices that are, in fact, not *quality* management. There are practices that simply automate and *institutionalize* information scrap and rework. There are practices of quality “assessment” that fail to provide a true measure of data *accuracy*. There are practices that in the name “data quality” actually create new data quality problems – *by design*! There are some who propose a “maturity model” for data quality that is only a taxonomy – it does not correlate to a true Quality Maturity Model as described by Philip Crosby.

The word “quality” is not a word to be used flip-pantly. To apply the word “quality” to an information or data quality “methodology” that implements “quality” practices to software that supports quality management or to the certification of data or processes as “quality-certified,” we must understand exactly what quality means and what it entails.

The purpose of this article is help those seeking to implement an information or data quality function understand the essential ingredients required to make the function an information “*quality*” management function and to make it an *effective* business management tool. The sidebar on page 38 describes the 10 essential ingredients required to put the word quality into your information or data quality initiative.



DATA
QUALITY



1. Understand information quality is a *business problem, not just a systems problem, and solve it as a business process, not just as a systems process.*

The goal of information quality management is *not* about improving what is in the data warehouse or even the source databases. The goal of information quality management is: **To increase business effectiveness by eliminating the costs of nonquality information and increasing the value of high quality information assets.** Information quality problems don't hurt systems – they cause business processes to fail. Unexpected data can cause an application to abnormally end, but the real problem is the business process failure and scrap and rework that results.

Information quality solutions are not system solutions – they are business solutions that encompass the business processes, applications, databases, the people who perform the work of the enterprise and the environment in which people work. Any so-called “solution” that addresses only part of this will be an incomplete solution and will be suboptimized at best.

Often systems personnel introduce information quality software to “solve” IQ problems by “analyzing” or “cleansing” the data. However, failure to identify the root causes of the nonquality data will invariably cause that “solution” to fail. There are many root causes of IQ problems, including: ill-defined processes, nonintuitive or misleading form design, untrained information producer personnel and harmful people performance measures as well as defective data design, disparately defined redundant databases and defective application design. Even putting robust system edits in source applications or implementing defect prevention IQ software at the source will not completely solve information quality problems if the cause is with the process, procedures, training or environment.

For example, an order department put the word “deceased” in a customer's name field when it found that the customer had passed away. This was the order department's code to not call. This was “fine” until marketing sent out letters to “Dear Deceased.” Data edits could prevent the name of the deceased from being inserted, but this would not solve the root cause: not understanding how updates affect downstream knowledge-workers.

2. Focus on the information customers and suppliers, not just the data.

An easy temptation when starting a quality program is to focus on the data. After all, that is what is defective. Isn't the goal of a quality program to improve the quality of the product?

Masaaki Imai, creator of the quality method Kaizen writes, “When speaking of ‘quality,’ one tends to think first in terms of product quality. Nothing could be further from the truth. In TQC [Total Quality Control], the first and foremost concern is with the quality of people ... A company able to build quality into its people is already halfway toward producing quality products.”^{1,2} Imai further states that only after the “humanware” of the business has been addressed in TQC, should the hardware and software aspects of business be considered.

Why do organizations need data anyway? Precisely because it is required for knowledge-workers to perform the processes of the business. The quality method must focus on the information “customers” to understand their quality requirements in order to

perform their processes efficiently and effectively. In W. Edwards Deming's 14 Points of Quality, Point 1 describes an obligation to the customer that never ceases because “the consumer is the most important part of the *production line*.”^{3,4} In information quality

management, this means “the obligation to the knowledge-workers” never ceases, as they are the most important part of the *information value chain*.⁵ Information customers include all persons within the enterprise who require information to perform their work and all external stakeholders, including end customers and shareholders, who depend on information. So important is the customer of quality management that it requires us to rethink how we treat information “customers” and the terms we use to refer to them.⁶

Armand Feigenbaum, creator of the Total Quality Control method confirms, “Quality is a *customer* determination, not an engineer's determination, not a marketing determination or a general management determination ... based upon the *customer's* actual experience with the product or service, measured against his or her *requirements* – stated or unstated, conscious or merely sensed, technically operational or entirely subjective.”⁷ Because knowledge-workers *require* information to perform their work, the definition of *information quality* – (consistently meeting *all* knowledge-worker and end-customer expectations through information and information services) – focuses on them.⁸

Information must meet the expectations of *all* knowledge-workers, not just those in one business area. The product price attribute is not a one-purpose fact required by those who take orders. It is required by accounts receivable to properly bill the customer, accounting to determine cost of sales, marketing staff to promote the products and product development to analyze product performance and product design. Information quality is not just fitness for a purpose. Information must be fit for *all* purposes to be “quality.”

Therefore, a true quality method must focus on information producers as important people in the information value chain. Information producers must know all downstream knowledge-worker customers and external end customers. Many information quality problems today come from data that meets only one narrow set of customer needs (*my* department). Information quality methods must equip information producers to capture information with quality that meets the needs of *all* knowledge-workers requiring it.

Managers of information producers must also have the resources to enable their staff to produce quality information. Then they must be held accountable for the quality of the information produced within their areas. This is one of the most important information stewardship roles, yet it is one of the most neglected.

3. Focus on *all* components of information, including *definition, content and presentation.*

As a business issue, information quality is not just about what is in the databases. It is about all forms of information – electronic, paper, oral, graphic and signage. Information quality methods must address quality of definition, content and presentation.⁹ Process failure occurs not just when the data in the database is missing or inaccurate. Process failure occurs when data has not been precisely defined and information producers create data

to mean what they think it should mean. Data definition is NOT documentation. "Data definition and information architecture are to data [content] what a product specification is to a manufactured product."¹⁰

The *Health Care Data Element Dictionary* contains a listing of health care industry-assigned data element names used in the nine HIPPA-related ASC X12N Health Care Implementation Guides. In that dictionary, the term *Payment Date* – (to illustrate how to use the document) is defined as "date of payment"¹¹ This not only violates the fundamental lexicographic rule of definition – do not define a term with the term itself – it is a blatant example of how poor data definition prevents quality of information content. What is the meaning of "date of payment"? Is it the date the payment was authorized, date the payment request was submitted, date the check was cut, date the check was put in the mail, date the payment received by the recipient or the date the payment entered into the system? With such a non-robust definition, any or all of those meanings could apply. A consultant who implements HIPPA in organizations defended this style saying that it provides flexibility. This is **not** flexibility. This is a nonquality data definition that **causes** nonquality data and communication problems.

Process failure also happens when information is presented ambiguously, whether in a report, on a computer screen or on paper. Labels on screens or reports may be misleading. A summary mortgage loan report had a column labeled "profit." When I asked the meaning, I was told this was the calculated profit that would be made if all the loans went to maturity. Because very few mortgage loans go to maturity, the report could be misleading. Whether management understood the real meaning of that data is irrelevant. The label "profit" and the definition were not consistent. Anyone not understanding *this* definition of profit could make some catastrophic decisions.

4. Measure data accuracy, not just validity.

I do not ever recall an information consumer from the business telling me that it is okay if the data is inaccurate as long as it is valid. Recently, I was taken back when a student in my *Information Quality Improvement* seminar told me she had a major revelation about IQ assessment when we discussed measuring accuracy. Coming from a firm that provides IQ assessment software, she had never thought about "accuracy" as a characteristic that had to be measured by comparing the data to the real-world object or a recording or observation of events. Unfortunately, some providers of IQ tools only focus on quality characteristics that can be measured or supported by their tools. Accuracy, timeliness, accessibility and presentation intuitiveness are quality characteristics required by knowledge-workers. These characteristics cannot be "assessed" by IQ analysis or assessment software.

We can learn from manufacturing quality

to understand how to perform information quality processes. Ishikawa, the early Japanese quality guru, describes the process of quality measurement for manufactured products. You take a sample of the product and measure its quality by comparing the physical characteristic of the product to the *data* – the product specifications.¹² Is the characteristic of the object within the specs? This, of course, assumes the specs accurately represent the expectations of the customers. To measure the quality of the manufactured goods, compare the object to the data.

How do you measure the information quality characteristic of accuracy? Because data is the electronic representation of real-world objects or events, you compare the data to the *real-world object* or an observation or recording of the *event*.¹³

A sound IQ assessment process must differentiate accuracy from validity assessments and report them properly. Validity is a measure of the degree of conformance to its domain of valid values *and* applicable business rules.¹⁴ Accuracy (to reality) is the degree to which data accurately reflects the real-world object or event being described.¹⁵

When one company measured validity of "marital status," they found zero percent invalid values. However, when they conducted an accuracy assessment, they found that **23.3 percent of those valid values were incorrect**. Only measuring validity but not accuracy may deceive your knowledge-workers into a false sense of security about their data "quality."

5. Implement information quality management processes, not just information quality software.

IQ software is a very

Ten Essential Ingredients of Information Quality Management

1. Understand information quality is a business problem, not just a systems problem; and solve it as a business process, not just as a systems process.
2. Focus on the information customers and suppliers, not just the data.
3. Focus on all components of information, including definition, content and presentation.
4. Implement information quality management processes, not just information quality software.
5. Measure data accuracy, not just validity.
6. Measure costs – not just percent – of nonquality information and business results of quality information.
7. Emphasize process improvement and preventive maintenance (Plan-Do-Check-Act), not just corrective maintenance (data cleansing).
8. Improve processes at the source, not just in downstream business areas.
9. Provide quality training to managers and information producers (who are their information customers and what do they need?).
10. Actively transform the culture, don't just implement activities.

important **part** of an information quality environment. With IQ products, we can accelerate many processes of IQ management such as:¹⁶

- Rule discovery and data analysis (data mining to discover business rule patterns, anomalies and potential duplicates)
- Validity assessment
- Data correction (some categories)
- Transformation control
- Enhancement (appending externally provided data to increase data value)
- Data defect prevention (some categories of accuracy and most categories of validity)

Even with many excellent IQ products on the market, it is a mistake to believe that the software is the silver bullet that solves *all* the information quality problems. Implementing the rigorous business rule and validation rules in the source-data capture will fail to guarantee quality data if data entry staff are entering data from a form created in another part of the business or if the information producers have volume quotas as performance measures.

A true quality methodology begins with understanding the business problems you are solving, implementing quality management processes and then selecting the appropriate IQ software tools. This approach increases your potential for an effective information quality environment.

One caution. Many IQ software providers are developing methodologies. Sometimes, those methodologies are designed around their products. If so, they may have limitations. For example, one provider gave me an early overview of their new IQ assessment service and method. It neglected to address accuracy assessment, which I pointed out to them. I believe their method still omits accuracy assessment, but their Web site does at least claim they assess data for violations to business rules (validity).

6. Measure costs – not just percent – of nonquality information and business results of quality information.

Manufacturing firms with quality programs know the important issue in quality is the cost of nonquality, not just the number of defect instances. Therefore, they measure and track the costs of (manufacturing) scrap and rework. The same principle is true in information quality. A sound IQ practice measures and tracks costs, not just percents of information scrap and rework. Knowing the costs of information scrap and rework enables you to measure the ROI of process improvements and helps you prioritize quality issues.

What management requests is the return on investment (ROI) of an IQ program because they consider the costs of information scrap and rework as “normal costs of doing business.” The same was true in the early days of manufacturing when quality was believed to be “inspection and scrap and rework.” As people came to understand the costs of nonquality, quality management shifted to a practice of “designing quality in” to eliminate the costs of waste.

Every sound quality management system confirms the truth that quality actually *reduces operating costs* and *increases productivity*. Deming’s Point 2 is “Adopt the new philosophy” of quality because “Reliable service reduces costs. Delays and mistakes raise costs.”¹⁷ Crosby, another of America’s quality gurus, correctly stated “quality is

not only free, it is an honest-to-everything profit maker. Every penny you don’t spend on doing things wrong, over, or instead becomes half a penny right on the bottom line.”¹⁸

The business case for any initiative is the cost to do it relative to the benefits to accomplish the enterprise mission. A sound IQ methodology will measure the costs of nonquality information and how it inhibits the enterprise objectives. From that baseline, we can measure the benefits when we improve the processes to increase IQ and reduce the costs of information scrap and rework.

7. Emphasize process improvement and preventive maintenance (Plan-Do-Check-Act), not just corrective maintenance (data cleansing). Without a concept of “customer” (See #2) and a defined and a habitually practiced improvement process, a method or practice cannot legitimately use the word “quality” management or methodology for several reasons.

The highest payoff for quality information is when we “design quality in” to the processes that produce it. Quality data that comes from clean-up initiatives incurs the added costs of the “information scrap and rework” of correcting data that could have been created and maintained by the processes in the first place. This does not account for the additional costs incurred from the process failure, workarounds, rework, customer alienation and lost business as well as missed opportunity that resulted from the defective data before it was corrected. Any value proposition for quality data by means of cleansing (corrective maintenance) alone is suboptimized.

Every valid quality system recognizes real quality comes by designing quality in. Deming’s Point 3 is “Cease dependence on inspection to achieve quality. Eliminate the need for inspection on a mass basis by building quality into the product in the first place.”¹⁹ The Kaizen philosophy assumes that our way of life, be it our working life, our social life or our home life, should focus on constant improvement efforts.²⁰

Does this mean we should not conduct data clean-up (corrective maintenance) initiatives? No! It *does* mean that we should address corrective maintenance of a set of information as a **one-time** activity. In parallel we **MUST** conduct a Plan-Do-Check-Act (PDCA) process improvement to identify the root cause of the nonquality and implement improvements to prevent recurrence. Solely addressing cleansing condemns the organization by “institutionalizing” defect production and data defect correction without solving the business problem.

One misconception companies have is that process improvements are expensive. Wrong. A process improvement can entail *relatively little cost*.²¹ Joseph Juran states, “Some managers hold to a mind-set that ‘higher quality costs more.’ This mind-set may be based on the outmoded belief that the way to improve quality is to increase inspection so that fewer defects escape to the customer.”²² Improvements can often be as simple and inexpensive as getting information producers together with their information customers and getting regular feedback.

8. Improve processes at the source, not just in downstream business areas. In *Kaizen*, quality must happen in *gemba*. “Gemba” means “the real place” in Japanese; and in quality improvement, that is where value is added.²³ In manufacturing, it is the shop floor. In the

service sector, *gemba* is where customers come into contact with services offered. In information quality, *gemba* is wherever information is gained or exchanged. That is where quality management must happen.

The data warehouse revolution brought attention to the problems of nonquality data. At the same time, most data warehouse methods that addressed quality improvement *actually exacerbated the quality problems* rather than solved them. These faulty “quality” techniques called for correcting data **not** in the source, but only in a staging area or in the warehouse after loading. This simply created a new information quality problem – data in the warehouse was *out of sync* (inconsistent) with the data in the source. Three problems are created by this so-called “data quality improvement” technique:

- Prevents drill through to audit the aggregations to the source data. Queries to the warehouse and queries to the source **cannot** be reconciled.
- Errors in the source can subsequently corrupt the warehouse data.
- Allows processes that use the still-defective source data to continue to fail.

Data warehouses developed this way created new silos of information and added complexity to the already disparate data environment.

A true IQ method will solve IQ problems in *gemba* where information first comes to be known and will maintain it throughout the value chain where that data is updated and maintained.

9. Provide quality training to managers and information producers (who their information customers are, quality requirements and how to meet them).

Training is so important in quality that it makes up two of Deming’s 14 Points. Point 6 states “institute training” of staff in how to perform their work with quality. “Management needs training to learn about the company, all the way from incoming material to customer.”²⁴ Management must understand the value chains of the enterprise to understand their information customers and suppliers. In information quality management, one must understand the information value chains to understand what processes depend on the information produced by his or her business area and what their *customer* requirements are for the information produced.

Training must be provided to the information producers who produce that information for the same reason. The producers also need training in the meaning of the data, the valid val-

ues and the business rules as well as *how* to produce the information to meet the needs of their “information customers.”

When management becomes convinced they need to emphasize information quality, they often try to get it using “slogans,” believing that

Checklist Ten Essential Ingredients of a Sound Information *Quality* Management Methodology

1. The IQ method holistically addresses the business process and not just the application systems and database aspects. It implements IQ processes as a management tool to increase business process effectiveness.
2. The IQ method has a strong focus on the customers of information to understand and meet their quality expectations, and a strong focus on the information producer and producer manager to provide them training and resources so both can accomplish their jobs successfully across the business value chains.
3. The IQ method addresses and assures quality of data definition and architecture design (as “information product specification”) and presentation quality as well as content.
4. The IQ method differentiates between validity and accuracy and never presents a measure of validity as if it represents accuracy. It emphasizes accuracy assessment for important data even if less frequent than validity assessment and with a smaller sample.
5. The IQ method addresses the proper perspective between defined processes for IQ management coupled with IQ software products to exploit their capabilities. The IQ processes have procedures for performing IQ functions not capable of being provided by the IQ software.
6. The IQ method measures the costs of nonquality information so management understands the *business* costs of the status quo and as the basis for measuring the benefits of process improvements.
7. The IQ method has an effective *process improvement* process, such as the Shewhart Cycle (Plan-Do-Check-Act), that analyzes root cause and defines improvements, implements and confirms improvements that eliminate recurrence of the defects and puts the process in control as a repeatable process. The method emphasizes training for everyone on how to conduct process improvements and encourages them to do so.
8. The IQ method emphasizes information quality improvements and control at the *source* processes that capture and maintain it.
9. The IQ method emphasizes training for managers and information producers to provide them the tools to better perform their jobs in creating and maintaining high-quality information.
10. The IQ method actively encourages and facilitates a cultural transformation, helping executive management understand its role in IQ, emphasizing management accountability for information quality and creating customer-supplier awareness among management and staff with “customer satisfaction” in knowledge-workers and end-customers as key performance measures.

if people know they should create quality information, they will. However, in order for people to create quality information, they have to be trained. They also need training in how to conduct their own process improvements and be empowered to improve processes where they find flaws.

10. Actively transform the culture, don't just implement activities. Many attempt to implement quality activities as one would a systems project. Some benefits can be gained, but the real benefits and improvements cannot be sustained over the long haul without a fundamental cultural transformation.

This cultural transformation is one of thinking and managing the business horizontally across the value chains. Management must help people understand the "customer/supplier" relationships internally in the information products people produce and depend on. Management must implement performance measures based on how well a business area's information products meet their internal information customers' needs and eliminate the downstream process failure and information scrap and rework caused by defective data.

Deming's Point 14 emphasizes that management must "take action to accomplish the transformation." He correctly states that management in authority will "struggle over every one of the [first] 13 points," but "will take pride in their adoption of the new philosophy and in their new responsibilities," and will "explain by seminars and other means to a critical mass of people in the company why change is necessary, **and that the change will involve everybody.**"²⁵

Philip Crosby has given us a Quality Management Maturity Grid whereby we can assess where our organization falls in its maturity of implementing quality as a "management tool."²⁶ Information quality management is a "management tool" for the Information Age. I have adapted Crosby's Management Grid as to how an organization may assess itself as to its own maturity. With that assessment, the organization then knows its next steps for maturity to the stage of certainty in which it knows "exactly why it does **not** have problems with information quality."²⁷

The word "quality" in a method or system of "quality management" has been defined in the manufacturing sector. The lessons learned in manufacturing quality demand key ingredients in "information" quality management. They are:

- Solve IQ as a business problem.
- Focus on the information customer as the subject of importance and the information producer as the agent of quality. (Remember most everyone is both an information customer and an information producer.)
- Address all components of information: definition, content and presentation.
- Implement information quality processes, not just software.
- Measure accuracy, not just validity.
- Measure costs of nonquality – the only measure that truly matters.
- Emphasize process improvement – not just "cleansing."
- Improve processes at the source.
- Provide training so people know "how" to make quality happen.
- Do not just implement IQ activities, but transform the culture to make quality a value system and mind-set in all products and a habit.

I fully believe that organizations that embark on this kind of

information quality environment will produce the same kind of economic revolution that Japan did in the '70s and '80s that changed the ground rules for competing in the manufacturing markets. Furthermore, I believe that those who ignore the IQ revolution will put their organizations at risk. DM

References:

1. Kaizen is a registered trademark of the Kaizen Institute. Kaizen means "continuous improvement ... that involves everyone—both managers and workers." Imai, Masaaki. *Gemba Kaizen*. New York: McGraw-Hill, 1976. p. 1.
2. Imai, Masaaki. *Kaizen: The Key to Japan's Competitive Success*. New York: McGraw-Hill, 1986. p. 43.
3. W. Edwards Deming is the American Quality consultant who went to Japan in 1950 and taught his quality method to the Japanese.
4. Deming, W. Edwards. *Out of the Crisis*. Cambridge, MA: MIT Center for Advanced Engineering Study, 1986. p. 26.
5. English, Larry P. *Improving Data Warehouse and Business Information Quality*. New York: John Wiley & Sons, 1999. p. 341.
6. See discussion of appropriate terms to replace the term "user." English, Larry P. "What's in a Name: A Mandate for the IT Industry," *DM Review* exclusive online content, September 2000, last accessed 7/13/2002 at <http://www.dmreview.com/master.cfm?NavID=55&EdID=2642>; Larry P. English, "The Term 'User' Has No Place in the Information Age," *DM Review*, May 2002, p. 38-41; and Larry P. English, "Reply to May Column Responses," *DM Review*, July 2002, p. 55.
7. Feigenbaum, Armand V. *Total Quality Control*. New York: McGraw-Hill, 1991. p. 7.
8. English, Larry P. Information Quality Improvement: Processes and Best Practices for Business Performance Excellence, 2002 Ed. Brentwood, TN: INFORMATION IMPACT International, Inc., 1992-2002. p. 1.2.
9. English. *Improving Data Warehouse and Business Information Quality*. p. 27.
10. *Ibid.*, p. 84.
11. Health Care Data Element Dictionary (HIPAA). Rockville, MD: Washington Publishing Co., 2000. p. 3. Available at http://www.wpc-edi.com/hipaa/HIPAA_40.asp. (Date of access: July 13, 2002.)
12. Ishikawa, Kaoru. *Guide to Quality Control*. Tokyo: Asian Productivity Organization, 1982. p. 109.
13. English. *Improving Data Warehouse and Business Information Quality*. p. 184.
14. *Ibid.*, p. 145.
15. *Ibid.*, p. 147.
16. For a discussion of categories of IQ software products, see English, *Improving Data Warehouse and Business Information Quality* chapter 10, "Information Quality Tools and Techniques," pp. 311-333. For an annotated listing of IQ products, their IQ function categories and links to the software provider Web site, visit <http://www.infoimpact.com>.
17. Deming, W. Edwards. *Quality, Productivity, and Competitive Position*. Cambridge, MA: MIT Center for Advanced Engineering Study, 1982. p. 21.
18. Crosby, Philip B. *Quality Is Free: The Art of Making Quality Certain*. New York: Penguin Group, 1979. p. 1.
19. Deming. *Out of the Crisis*. p. 23.
20. Imai. *Gemba Kaizen*. p. 1.
21. *Ibid.*
22. Juran, Joseph M. and A. Blanton Godfrey. *Juran's Quality Handbook*, ed. New York: McGraw-Hill, p. 5.11.
23. Imai. *Gemba Kaizen*. p. 13.
24. Deming. *Out of the Crisis*. p. 52.
25. *Ibid.*, p. 86.
26. Crosby. *Quality Is Free*. pp. 21-119.
27. English. *Improving Data Warehouse and Business Information Quality*. pp. 427-437.

Larry English is the president and principal of INFORMATION IMPACT International, Inc. English is an internationally recognized speaker, teacher, consultant and author and can be reached at larry.english@infoimpact.com or through his Web site at www.infoimpact.com.