

7 Deadly Misconceptions

about

Information Quality

By
Larry P. English

Seven DEADLY Misconceptions about Information Quality

by Larry P. English

©1999, INFORMATION IMPACT International, Inc.

This white paper describes some fatal misconceptions about information quality that can cause information quality initiatives to fail—or to appear to succeed but fail to achieve positive business results. If an organization misunderstands the fundamentals of information quality improvement, it will fall into the same trap of every other “methodology” silver bullet.

There are seven potentially fatal misconceptions about information quality that can hamper an information quality initiative. Worse yet, if these misconceptions are strongly held, they will hamper business effectiveness (best case) or result in business failure. The seven deadly misconceptions are:

1. “Information quality is data cleansing”
2. “Information quality is data assessment”
3. “Conformance to business rules is the same as data accuracy”
4. “‘Information quality is data accuracy’; and its counterpoint: ‘Information quality is ‘fitness for purpose’”
5. “‘Information quality problems are caused by information producers’; and its corollary: ‘Information quality is produced by an information quality group’”
6. “Information quality problems can be edited out by implementing business rules”
7. “Information quality is too expensive”

Misconception 1: “Information quality is data cleansing”

Some think that by “cleansing” or “correcting” data they are improving information quality. Not true. While data cleansing and correction does “improve” information *product* quality, it is merely information “scrap and rework.” Like manufacturing scrap and rework, data cleansing is merely rework to correct defects that would not be there if the processes worked properly. To be sure data cleansing is required for any data warehouse or conversion project to succeed. If data in a data warehouse is nonquality, the warehouse will fail.

Data cleansing and correction are, simply put, part of the *costs* of nonquality data. Every hour consumed and dollar spent correcting data is an hour or dollar that cannot be spent doing something that adds value. Information quality is quality data produced at the source. Information quality improvement databases are designed properly and the processes are defined and operating properly.

Information quality improves processes to *prevent* defective data from being created. Data cleansing attacks the symptoms of a problem. It fixes the *results* of faulty processes. Information quality attacks the root causes of data defects and eliminates the *causes* of

nonquality information. A truly effective data “cleansing” function is one that works itself out of a job! It will transform itself from an information scrap and rework function to a function that facilitates data defect prevention.

Misconception 2: “Information quality is data assessment”

Another common misconception is that information quality is data assessment. No, again. Data audit, analysis or assessment is simply inspection. The immediate goal of assessment is to discover defects. While some data is so important it must have regular audits and controls in place, data assessment is a cost activity that does not, in and of itself, add value. Assessment of data quality has value when it is used to raise awareness of process failure *and* results in process improvements that eliminate the cause of defective data.

The *ultimate* goal of information assessment must be to assure processes are creating and maintaining quality that consistently meets *all* information customers’ requirements. Discovery of unsatisfactory information quality must lead to information *process* improvement and control.

Information quality minimizes data assessment because information quality is designed in to the processes and controlled during information production. An effective information quality function uses data assessment as a tool to improve the processes that create, maintain and deliver information.

Misconception 3: “Conformance to business rules is the same as data accuracy”

There is a temptation to equate the fact that data that conforms to business rule tests applied in automated data analysis means the data is accurate. Data that conforms to the business rules simply means the data has *validity*. That is, it is a valid value according to the defined rules. The reality is that many data errors are valid values that conform to all specified rules, yet are incorrect. One bank discovered that it had 1700 customers who had a birth date of November 11, 1911—a proportion way out of the normal frequency for their customer population. The data was valid, i.e., it fell within the range of valid birth dates for its customers. However, virtually every one of those values was inaccurate. The cause? The edit rules required a value for birth date. When the information producers did not know it, they simply entered “111111”—the fastest valid date they could enter!

Automated data quality assessment software that tests data only for valid values and conformance to other reasonability and calculation rules must report the data as having *validity* and clearly indicate this does not imply *accuracy*. Confusing validity or conformance to business rules with accuracy can create a false sense of security that the processes are working properly, when in fact they may not be. The results of decisions made based on valid but inaccurate data can be just as devastating as the results of decisions made on invalid data values.

Data is a representation of real-world objects or events. Data *accuracy* means that the data *correctly* represents the real-world object or event it characterizes. Accuracy means the facts are correct.

Misconception 4: “Information quality is data accuracy”; and its counterpoint: “Information quality is ‘fitness for purpose’”

Isn't it something of a heresy to say that it is a misconception to equate data accuracy and information quality? After all isn't one of information quality improvement's goals to improve the accuracy of information. The answer to that question is yes. But it is wrong to think that just because data is accurate it has quality. Quality exists only when a product is being used. Data in a database that is one hundred percent accurate is not quality if it is not used (or able to be used) to accomplish the work and mission of the enterprise. More on this in a moment.

Likewise, it is a misconception to equate information quality with fitness for purpose. The reason is simple. Data is not a resource that supports only *one* process or purpose. A notable example is the insurance company that discovered 80 percent of its claims were paid for a diagnosis of “broken leg.” No, they were not in a rough neighborhood. The claims processors were paid for how many claims they processed. So they let the system default to “broken leg” for the diagnosis code. The process to pay a claim did not require that the diagnosis code be correct—only that it be a valid code. This data actually had “validity”—the values were valid values, that is, the data conformed to at least one of the business rules.

The quality problem did not surface until the data was loaded into a data warehouse for the actuary to analyze risk. The data in this example was in fact “fit for purpose” to pay a claim. But it was totally worthless for the risk analysis process.

So what is information quality? Information quality is fitness for *all* purposes in the enterprise processes that require it. It is this phenomenon of fit for “my” purpose that is the curse of every enterprise-wide data warehouse project and every data conversion project.

The real tragedy of one business area creating data to meet only its needs, however, is that the data cannot be used by processes outside their business area. It forces downstream knowledge workers to have to re-acquire the same data, or engage in information scrap and rework to clean up and correct the data problems caused by the original source processes.

Now back to accuracy. Accuracy is only one *characteristic* of quality, just as validity or conformance to business rules is one characteristic of quality. These characteristics are some of the information quality characteristics categorized as *inherent* characteristics.

Fitness for purpose is the characteristic of *usefulness* of data for a specific requirement, that is, data is useful for me in performing my work. Usefulness, timeliness and presentation clarity are characteristics of what is classified as *pragmatic* quality. Those are the characteristics that facilitate knowledge workers to do their jobs efficiently and effectively.

The truth is that data requires both accuracy and fitness for all purposes, along with other required characteristics expected by its information customers to be considered quality information. This is data that “consistently meets knowledge workers and end-customers expectations.”

Misconception 5: “Information quality problems are caused by information producers”; and its corollary: “Information quality is produced by an information quality group”

These related misconceptions misunderstand the root causes and solutions for information quality problems. The belief that information quality problems are caused by the people who create the data is based on a need to find “blame.” If information quality problems exist, the *processes* are broken, not the *people* who perform the processes. In the example of the “broken leg” problem cited above, the claims processors who are doing exactly what they were trained and rewarded for—paying claims as fast as possible. They were merely doing what they understood their job to be.

Information quality means analyzing the symptoms (the defective data) and the processes that produced the defective data to discover what has caused the defects. Root cause analysis gets behind the *immediate* cause to discover the *originating* cause. Causes may be inadequate training, lack of understanding of downstream knowledge workers and the uses made of the data or unclear or incomplete process procedures. In many cases the cause is confusing “productivity” with speed of work, and measuring the wrong things. In the “broken leg” problem, management had created incentives for how many claims a processor handled per day without considering the impact of nonquality information on any downstream processes. The real diagnosis for this was “broken process!”

Information quality improvement requires a non-judgmental, blame-free environment. “Fault finding” only creates fear and stymies creative change. It leads people to cover up “problems” and not be open to exploring process improvements.

Likewise, it is a misconception to think that the information quality team is the “savior” and “solver” of information quality problems. Information quality problems are the result of broken processes from the top of the enterprise to the bottom and from the front line processes to the shop floor processes to the back office processes. One information quality team or organization cannot physically “manage” all information quality. Rather, a fundamental principle of kaizen is that everyone in the enterprise must take responsibility for their processes and their information products.

Information quality cannot be delegated. Each and every person in the enterprise must assume accountability for their role with information, whether they produce it, transcribe it or use it, or whether they design and define it, or build applications to capture or retrieve and present it.

So, what is the role of the information quality organization? First, it must sensitize the enterprise to the problems caused by nonquality information. Then it must define processes for measuring information product specification (data definition and information architecture) and data (content) quality. It must define processes for improving information product quality (cleansing and reengineering) and for information process quality. The information quality team must provide for education and for facilitating the enterprise in the improvement of processes where information quality does not meet acceptable quality standards. Information producers generally know the problems with their processes. Give them a process to improve, empower them and provide facilitation where necessary. Then measure the cost savings and increased customer satisfaction that results from improved information quality.

Misconception 6: Information quality problems can be edited out by implementing business rules

There is a major movement today that addresses defining business rules. Business rules are business policies that govern business actions, and as such, may provide integrity rules for data. The temptation is great to think that once one has defined and implemented business rules they have created process quality. Not true. While-well implemented business rules provide important editing and validation of data, if not accompanied with process quality principles, the very rules that are meant to improve information quality can actually *guarantee* just the opposite. Consider the bank that discovered over two dozen of its Customers had the same social security number. Root-cause analysis revealed that an edit routine required a valid nine-digit SSN number when creating a customer record. But if the customer did not know their social security number, one information producer simply entered her own, because the application would not let her create a record without a nine-digit number. Valid social security number? Yes. Quality data? No! The implemented business rule actually “edited in” nonquality data.

There are two important requirements to move from “implemented business rules” to “information quality improvement.” The first is to implement business rules properly. Allow “unknown” or null values when the information producers do not know the data. Forcing a value to create a record when that value is not known creates nonquality. Note that it is *not* acceptable to create “dummy” values like “999-99-9999” or “000-00-0000” to represent the absence of a social security or social insurance number. To do so requires applications to have logic to exclude such values from being used. Without this “exception” logic, processes using these “dummy” values will fail, providing invalid tax reporting, for example. Implement integrity rules at the right place. When creating knowledge about customers, it may not be possible to capture all facts about them, although the quality principle is to capture all possible data at the point the knowledge is knowable. Provide reasonability editing and validation at the creation of each attribute, including duplicate record matching. But, when creating a loan for a customer, certain facts such as social security number *must* be known. It is the process of “create loan” that must validate and verify that all required customer data exists and is correct. Implement edits within processes that use the data to assure process integrity. Finally, recognize the limitations of automated business rules. They define the required business policies, but cannot guarantee data accuracy through application edits. Implement business rules from an “error proofing” perspective (Juran calls this “fool proofing”) to prevent inadvertent human error but recognize that business rules in and of themselves cannot prevent inaccurate values.

The second requirement to move from business rule definition as an academic exercise to information quality improvement is to provide adequate training to information producers and provide clear business procedures. The people who create data must understand: (1) who are the information customers who use the data; (2) what are their quality requirements; (3) how is the data used and: (4) what are the costs of nonquality data. Without adequate training of the information producers and without

effective procedures, all the implemented business rules in the world will not produce quality data.

Misconception 7: Information quality is too expensive

The most fatal misconception is that it costs money to produce quality information. Just the opposite is true however. Information quality proponents around the world are being asked for a cost justification for changing the status quo—that is, spending money to “improve” the current processes. The question is based on a perception that the current processes that produce information must be working properly. After all, we are conducting business successfully, and we are making money. While this is a fair question to ask, it is the *wrong* question. The real cost justification question is, “Can we afford the costs of information scrap and rework?”

The tragedy in this is that management has accepted the costs of poor quality information as a “normal” cost of doing business. In fact, I routinely find, as I have recently with telecom companies, insurance companies, financial companies, and manufacturing companies, that top management is generally unaware of the real costs of nonquality data. Management removed from the actual operations may not see the costs of nonquality information. Why do organizations accept the costs of knowledge workers hunting for information, having to correct inaccurate data, sending inaccurate bills, requiring customers to have to change their addresses multiple times—because they exist redundantly in line-of-business databases—as normal business costs? Because of a misconception that this rework is not really hurting the business.

However, numerous information quality-cost analyses I have conducted illustrate that the direct costs of nonquality information in the typical organization claims 15 to 25 percent of its revenue or operating budget. The costs of rework, workarounds, data correction and cleanup, creating and maintaining proprietary databases because of inaccessibility to or nonquality data in production databases, multiple data handling of data in redundant databases, etc., etc. have an incredible, but often transparent, toll on the bottom line.

Is my experience unique? Information scrap and rework is to the Information Age what manufacturing scrap and rework was to the Industrial Age. Philip Crosby (*Quality is Free*) found the costs of manufacturing scrap and rework to be 15-20 percent of revenue. Juran (*Juran on Planning*) finds the costs of poor quality to be from 20-40% of sales. W. Edwards Deming (*Out of the Crisis*) cites Feigenbaum’s estimate that “from 15-40 per cent of the manufacturer’s costs of almost any American product . . . is for waste embedded in it.” The authors of the BBC video *Quality in Practice* cite the costs of quality in the typical manufacturing company to be around 20% of sales, while those of the typical service company are around 30% of sales. In the Information Age, nonquality information contributes to nonquality products and services.

Management can—and will—understand that the costs of nonquality information are unacceptable, when information professionals help management quantify the costs of nonquality information in tangible, bottom-line costs to the business. Remember that American management accepted the costs of manufacturing scrap and rework until the

Japanese illustrated these costs are not necessary and through continuous process improvement eliminated the costs of scrap and rework.

Management will likewise accept the costs of information scrap and rework *until* the competition eliminates their nonquality data, thereby reducing its costs of information scrap and rework, increasing its product and service quality and increasing its customer satisfaction, resulting in increased market share.

The bottom line is that quality information increases the bottom line—both in reduced costs of conducting business, and in increased opportunities as a result of accurate and managed knowledge about customers, products and services, sales, and other important things.

Conclusion

What then is information quality? It is quality in all characteristics of information, such as completeness, accuracy, timeliness, clarity of presentation that “consistently meets knowledge worker and end-customer expectations” to meet their objectives. The *process* of information quality improvement is one of continuous process improvement of any and all processes, to eliminate the causes of defective data. The *purpose* is to reduce costs of information scrap and rework and process failure, to increase customer and employee satisfaction and to increase business opportunity and profits.

What do you think? Send your comments to Larry.English@infoimpact.com or through his Web site at www.infoimpact.com.

About the Author:

Larry P. English is president and principal of INFORMATION IMPACT International, Inc., Brentwood, TN. Mr. English is an internationally recognized speaker, teacher, consultant, and author in information management and information quality improvement. DAMA awarded him the 1998 “Individual Achievement Award” for his contributions to the field of information resource management. He chairs The Information and Data Quality Conferences to be held in London, Oct. 11-13, 1999 and in New Orleans, Oct. 31-Nov. 4, 1999. Mr. English is the author of *Improving Data Warehouse and Business Information Quality: Methods for Reducing Costs and Increasing Profits*. Preview Chapter One at www.infoimpact.com.