

## Expert Insight: *Managing SCM Performance*

By Kate Vitasek

April 7, 2008

### *How Good is Your Supply Chain Data Quality?*

*The Key Is Understanding Data Quality from the User's Point Of View; Introducing the Data Quality Index*

You might find it odd that I start out a Blog dedicated to Performance Management asking the question, "How's your data quality?" But it is information and the flow of data that is at the heart of today's extended supply chains. It is data that has, in part, helped to fuel impressive productivity gains that most industries have enjoyed over the past few years. It is data that is fundamental to all performance metrics. And it is data that workers and executives alike have come to rely on to make decisions. Yet few companies treat data as the valuable asset that it is and few acknowledge the risk that poor data quality presents.

How big is the problem? A Gartner survey of 600 executives in November 2005 found that more than 25% of critical data within business is inaccurate or incomplete. Yet, we are asking our employees to trust the data as accurate when their experience tells them it is not.

I also quizzed an expert to see just how bad data quality is. According to Horacio Woolcott, Chief Executive Officer for PartsRiver (a data quality service company) "Our clients tend to overestimate the quality of their data and are shocked when we do an analysis and find out how poor the data actually is. Few have any metrics in place to measure critical data elements and, if they do measure, they seldom look at how well the data meets the needs of the user."

How do companies assess data quality? That's the problem, many do not. Few have a formal method for tracking data quality; they base their assessment on gut feel or may have looked at it as part of a major IT project. Most, however, do not know if they even have a problem. It is time that data quality, across the entire extended supply chain, gets the respect it deserves.

Understanding data quality through performance metrics is critical to developing a comprehensive plan to improve data quality and then maintain it. The old standard; "you can't improve what you don't measure" is as true with data quality as with any other process or operation.

The first step is to measure data quality in an objective way – from the user's perspective. Unfortunately, many companies do not make the effort to measure the quality of their data in any objective or quantitative way, believing it is too difficult or overwhelming a task. The task can be simplified; let's look at this from the user's perspective.

All users would agree for data to be usable, it must be accurate, consistent, complete and it must be valid. If any one of these data quality characteristics is not met, the information in that data element may not be usable. To assess user level data quality, we can focus on these four metrics.

- % Accurate: the percent of correct data elements. Elements without errors in any field.

- % Consistent: the percent of data elements that is consistent (identical data attributes) across data bases without duplication.
- % Complete: the percent of data elements that have values in required data attributes.
- % Valid: the percent of data elements with data attributes that meet the field requirements.

Each of these simple ratio metrics can be calculated from a sampling of data in a data base, for example, part master data or vendor master data.

To make the metrics more meaningful, the four key measures can be combined into a Data Quality Index that provides a realistic view of the impact of a company's data quality on the user. The approach is similar to the traditional approach manufacturers have used for years to measure First Pass Yield in a production factory where each production process is measured and total "yield" or fallout of the entire process is calculated as an index. Some of you will also recognize the similarity to the Perfect Order Index, applying the concept to data quality is reasonable.

The Data Quality index is calculated by multiplying the values of the four metrics (% Accurate, % Consistent % Complete, % .Valid). For example, if a company achieves 90% data accuracy, 90% data consistency, 90% data completeness, and a 90% data validity and viewed each metric as a separate element, the company may think that they are doing well, because each metric is at 90%.

However, if you combine the four metrics, the Data Quality Index is 65%. Or 35% of the data has an error in accuracy, consistency, completeness, or validity that may render the data unusable or unreliable. Therefore the users' experience is that data they use is correct only 65% of the time, hardly a passing grade.

From the users' perspective, the Data Quality Index is a good indication of how well data will support them in performing their work. This simple index can also concisely communicate the state of data quality to senior management, and provide a basis for comparative assessment of data quality improvement efforts.

The Data Quality Index is one way to quantify the level of "trust" the user has in the data. It can also be considered a "believability index," providing an indication of how the data is regarded as true and credible. Data that can be trusted and believed should be the ultimate goal.

Understandably, if the average company is currently at a 25% error rate as Gartner survey suggests, getting there will require substantial effort and commitment. It is clear that companies who take up the challenge to address the quality of their data across their extended supply chain will establish a competitive advantage over those who do not.

Our next installment on Measuring Data Quality will borrow from a Six Sigma production measurement, DPPM, and extend it to measuring data quality; I promise no boring statistics lecture.