



by jonathangeiger

Business Responsibilities for Managing Data

In last month's column, I described the data stewardship responsibilities in planning, acquiring, managing, disseminating and disposing of data. The stewardship role is shared, with the technology group typically responsible for the data handling processes (e.g., systems and databases) and the business groups responsible for establishing the policies and quality expectations and for providing the content. Quality management programs apply the continuous improvement cycle, root cause analysis and other techniques, and these same principles should be applied to managing the data assets.

The continuous improvement process, also known as the Shewhart Cycle, was developed by Dr. Walter A. Shewhart and popularized by Dr. W. Edwards Deming. The cycle recognizes the importance of planning a task, performing it, measuring the results, taking actions based on the measurements and then repeating the cycle. The continuous improvement process is sometimes called the plan-do-check-act cycle.

The cycle recognizes the importance of planning a task, performing it, measuring the results, taking actions based on the measurements and then repeating the cycle.

Proactive programs begin with the *plan* step. Within the planning step, the flow of the process is mapped out along with appropriate measurement points. The process may be manual or automated. An example of a manual process is the interaction that a salesperson may have with a customer that results in capturing information or any workflow process within the organization. Measurements may address the time spent, the completeness of the data or the accuracy of the data. An example of an automated process is extract, transform and load (ETL) processing. The process description includes the mapping specifications, and audit and control points provide common metrics.

The execution of the process is performed within the *do* step. The process is carried out in accordance with the specifications defined within the planning step. When programs are developed without good requirements and specifications, this is the first step, which is not a recommended approach.


The *check* phase provides a realistic assessment of the success of the process. The measurements need to be monitored both as single events to track performance relative to

the targets and over time to track the consistency of the performance. If the process is performing consistently and meeting the targets, then no action is needed in the next step. Otherwise, the reasons for the deviations need to be analyzed to determine the appropriate corrective actions. A common approach is to perform root cause analysis. With root cause analysis, I don't stop with the symptom (such as the detection of a data error). I delve more deeply to try to understand why the error occurred; for example, detecting that critical data is missing. The missing data is the symptom. I may investigate a little and discover that the data was not required by the source system. If I stopped there, I might recommend a maintenance fix to the system. A deeper investigation may reveal that salespeople were simply not capturing the data. The more comprehensive corrective action would address the business processes and possibly the incentive plans if the data is deemed to be important.

As part of a data warehousing project, the condition of the source system data should be analyzed. If the system is operating within an overall quality management framework, this step is simplified because metrics already exist and corrective actions are taken as needed. More commonly, the quality of the source system data is formally reviewed as part of the ETL development process. The technique is commonly called data profiling, and I will explore that topic in my next column. Data profiling looks at the results of the source system activities reactively and provides information on the quality of the data.

The last step is *act*. Within this step, corrective actions are determined so that they can be incorporated into an improved process. This would include the incorporation of program fixes and business process improvements. Once the improvements are made, the cycle starts over again with the plan.

Managing the quality of data needs to be approached by applying traditional quality management techniques. The continuous improvement process is a powerful tool for ensuring quality. It starts with work that is well planned. The planning process includes both a description of the tasks to be performed and the measurements that will indicate the quality of the process and the results. Each time a process is performed, the measurements are taken and, based on the results, improvement actions may be identified and pursued.

Next month, I will delve deeper into the check step, with particular emphasis on the data profiling activities that need to be performed as part of a data warehousing initiative. 

Jonathan G. Geiger is executive vice president at Intelligent Solutions, Inc. He may be reached at jgeiger@intelsols.com.