Want to Maximize Data Quality? Make Data an Enterprise Asset

Claudia would like to thank Jonathan Geiger for his contribution to this month's column.

Maximizing data quality continues to be a significant challenge for anyone building or maintaining a data warehouse. Although data quality issues arise anywhere, they are rarely caused by the process of moving data from the source systems through the data warehouse and on into the data marts. Typically, data quality problems occur in the source systems, often originating in the business processes themselves.

If the problem is not in the extraction, transformation and load processes, what can you and your organization do to address data quality? There are three prerequisites for maximizing data warehouse data quality:

First, *data should be accepted as an asset.* An organization must consider its data to be as important as any other asset (e.g., human resources, facilities or cash). Data quality needs to be part of everyone's job. In fact, it should be included in each job description. It should be a parameter of performance evaluations and incentive packages. Many reward systems are focused on completeness (i.e., fill in the form), rather than accuracy (i.e., fill in the form with correct information). As long as this is the case, data quality will continue to suffer and remain a serious problem.

Second, *employees should be assigned responsibility for data.* The chief financial officer and controller are responsible for establishing and enforcing policies concerning the acquisition, use and disposition of the monetary assets of most corporations. That is, they establish and maintain rules which govern each employee's authority with respect to processing corporate funds. Likewise, people in the organization must be assigned responsibility for establishing and enforcing policies concerning data classification, quality, standards and processing.

When no one steps up to the plate to define the data quality parameters, the data warehouse team spends a tremendous amount of time obtaining opinions, soliciting solutions, facilitating a decision or, if one is not obtainable, creating workarounds. This time could better be spent developing the next iteration! Until the organization assigns this "stewardship" responsibility, the data warehouse team will continue to spend much of its time trying to understand existing data and trying to resolve the differences in the business meanings. Only after achieving this understanding or resolution can they tackle creating the transformation programs.

Unfortunately, this role is rarely filled. Major reasons for the failure to fill this role include: not recognizing data as an asset, political or cultural considerations (e.g., who should be responsible for customer data?), the difficulty involved and other priorities. Often, the other priorities involve fixing data problems (which, of course, wouldn't exist were this function established!).

Third, *data should be modeled like other assets.* Just as human resources are modeled via organization charts, facilities are modeled via blueprints and accounting is modeled via a chart of accounts, data requires modeling via the business or enterprise data model.

When this model exists, creating the data warehouse model is a fairly straightforward and simple task. It involves transforming the business data model into one appropriate for the data warehouse through a series of logical design steps that can often be performed very quickly – usually in less than two weeks for each data warehouse iteration. When the business data model doesn't exist, the data warehouse modeler needs to develop a portion of this model before proceeding to the data warehouse model. This step can also be accomplished in a reasonably short amount of time.

> **Do everything possible to influence your business and technical community to treat data as an important corporate asset.**

We recommend that you do everything possible to influence your business and technical community to treat data as an important corporate asset. That said, data warehouse practitioners have two choices when it comes to handling low quality data. They can either promulgate the errors or fix them, which creates data in the warehouse that does not match the system of record.

Let's address whether the data warehouse should match the system of record. Should we build a data warehouse that follows the philosophy of "garbage in – garbage out" even though we recognize and could fix certain errors? Is consistency with the source system's data more important than accuracy?

Proponents of being consistent with the system of record argue that the company uses this information in its day-to-day decisions. Therefore, it should be good enough for the

warehouse and strategic decision making. Additionally, the data warehouse team or the business community using the data won't need to reconcile differences between its information and the information that exists on current reports.

Advocates of correcting the errors argue that we should provide the best information possible and that if we can fix a problem, we should. With proper documentation and meta data, the reconciliation process should not be as difficult or onerous. Further, this documentation should be sent back to operations so they may correct the problem at the source.

Once data is recognized as an important asset, the decision concerning the quality of the data in the warehouse is straightforward. The person or group responsible for each set of data would define the quality expectations. The data warehouse team would provide information concerning the cost of providing the desired level of accuracy, and the steward would ultimately make the decision to achieve or not achieve that level of quality.

While partnerships are struck between the business and technical communities for the design of the BI environment, the technical community is typically in the leadership role with respect to data quality. This should be reversed. The business community should assume the leadership role and must recognize that part of their role entails managing data as a corporate resource. Until then, you need to continue working along a dual front — developing the warehouse in partnership with your business community and soliciting their support in recognizing data as a corporate asset. DM

Claudia Imhoff is the president of Intelligent Solutions, Inc. Imhoff is a popular speaker and internationally recognized expert on the Corporate Information Factory, business intelligence and CRM. She has coauthored four books and more than 40 articles on these topics. She may be reached at cimhoff@intelsols.com.

Jonathan Geiger is an executive vice president at Intelligent Solutions. His 30-year career includes data and repository management, data warehousing, project management, systems development and support, planning, training and quality assurance. He may be reached at JGeiger@IntelSols.com.