



You Can be Sure if It's . . . Been Certified, Part 1

Jonathan would like to thank Richard Blahunka, data warehouse architect at Entergy Corporation based in New Orleans, for his contributions to this column.

The Good Housekeeping Seal was instituted in 1909 and provides consumers with confidence when they buy products that have this seal. Underwriters Laboratories, Inc. (UL) has been providing another widely recognized assurance since 1903 to consumers buying products tested by this institution. With regulations such as Sarbanes-Oxley, which impose financial and prison penalties for inappropriate use of information, it is becoming increasingly important to not only have quality data, but to be in a position to defend that data. In this and the next two columns, we describe an approach for providing business users with confidence in the quality of their data warehouse information.

Your company can take measures to provide a certification process. The commercial certification processes mentioned earlier have several things in common: they apply standards appropriate for the product and its usage, the group providing the certification is recognized as being capable to evaluate and certify the quality, the certification process is repeatable, and the consumers recognize and value the meaning of the certification.

Ask yourself a question - is all of your data worth the time and effort to ensure that it meets the quality standards? The answer, which may surprise you, is no! When you're dealing with the company's financial data, data completeness and accuracy (two aspects of data quality) are critical. When you're dealing with customer data, its accuracy and completeness are still very important, but there may be components of that data for which the accuracy and completeness are not as critical. For example, although a customer's age may help cluster customers for specific marketing campaigns, you may be willing to accept a certain level of incompleteness (if a customer is unwilling to provide his or her age) or a certain level of inaccuracy (by accepting information provided by a customer without verification).

Certification Levels

Just as different standards apply for different products and uses, the data quality requirements also vary. Therefore, it is appropriate to consider different levels of data quality certification. Let's consider some potential levels.

Bronze Level: Data at the bronze level is data that can be verified to have been processed properly throughout the data warehousing environment. This is the minimum that the business community should expect from the data warehousing environment. This level of certification does not ensure that the data is accurate or complete. It merely ensures that the processing performed in migrating the data from the source systems to the data warehouse and data marts and into end-user queries and reports was done properly. Measuring compliance requires a set of business rules for moving the data through the data warehouse environment. Measurements can be incorporated into the audit and control processes, and these can be tracked and reported.

▶ Is all of your data worth the time and effort to ensure that it meets the quality standards?

Silver Level: Data at the silver level goes a step further. Data at this level can be certified as having been processed correctly not only by the data warehousing environment, but also by the operational system environment. When the data can be certified as being processed correctly by the operational systems, we know that everything possible within the control of the operational systems and the data warehousing environment group has been performed properly. It means that the business data quality expecta-

tions have been established, that reasonable data entry quality checks have been performed and that data that has been accepted has passed these checks. Metrics for the silver level are conceptually similar to the bronze level, except they must include all of the operational systems as well.

Gold Level: Data at the gold level goes yet another step further. The quality of this data can be traced all the way back to its initial entry into the company. It means that in addition to the systems environment, the business processes can be verified to have been performed in a way that ensures the data meets quality expectations upon capture and that the data was properly handled throughout its life cycle. Executives using data at this level for regulatory and share-

holder reporting can be confident of its accuracy and completeness. For gold level certification, the business processes also need to be documented with measurement points. The collected metrics can then be monitored, and this would further support the company's ability to demonstrate that its business controls are being observed, as required by Sarbanes-Oxley.

Uncertified Data: Any data that either has not undergone the certification process or does not at least conform to the bronze level is uncertified. This includes data that has been manipulated subsequent to its provision by the data warehousing environment. For example, if a business user uses the results of a query in a spreadsheet to produce a report, the data is uncertified

unless that portion of the processing also undergoes the certification review.

This column introduces the concept of data certification. In next month's column, we will describe the Data Quality Certification Oversight Committee. In the November column, we will describe the certification process. 

Jonathan G. Geiger, executive vice president at Intelligent Solutions, Inc., has been involved in many Corporate Information Factory and customer relationship management projects within the utility, telecommunications, manufacturing, education, chemical, financial and retail industries. He has authored or co-authored numerous articles and three books, presents frequently at national and international conferences, and teaches several public seminars. Geiger may be reached at JGeiger@IntelSols.com.