The Data Warehousing Disconnect

BY GIANLUCA GARBELLOTTO

In the pantheon of famous last words, which includes Julius Caesar asking, "Et tu, Brute?" and Civil War General John Sedgwick claiming, "They couldn't hit an elephant at this distance" just before being killed by a Confederate sharpshooter, we need to add a new entry: "A data warehouse will solve all of our problems,"—former IT director.

Maintaining an effective connection between detailed data (entries, documents, transactions) and end reporting is the ongoing quest of any management accountant. Enterprise resource planning (ERP) systems or implementing data warehouses often have been touted as the "silver bullet" for an organization's information problems, but these not-so-easy "easy" solutions fall short of solving an organization's real problems.

The promise of ERP systems to provide a centralized, all-inclusive solution for all data and processes within an organization aims to solve the disconnect at the root. The failure to live up to this promise is evident from a number of indicators present in most, if not all, information systems.

◆ Different information systems are still used in subsidiaries and peripheral units.

◆ Older applications are used for specific needs that aren't covered as effectively by the main ERP system.

◆ The ever-present proliferation of official and unofficial spreadsheets used to provide different views of data for specific purposes and reconciliation continues unabated.

Indirect proof of the failure of the ERP promise also lies in the widespread use of ETL (Extract, Transform, Load) and related business intelligence applications. These are based on the concept of a central data repository to which all corporate data, whether residing in the main system or in peripheral, purpose-specific applications, is transferred for consolidation, analysis, and end-reporting purposes.

Do these solutions really add value to the business reporting supply chain, or do they create more issues than they actually solve? By imposing an intermediary data warehouse between the source data and the end reporting, these applications also impose predetermined rules and dimensions to drive the data "transformation" (the "T" in "ETL"). Does this cause an even greater disconnect between source data and end reporting? Are there alternative approaches that can meet the same goal with additional overall advantages?

The answer to all these questions is "Yes." A direct link between detailed data and end reporting is made possible by a standards-based approach based on XBRL in both of its "flavors": XBRL Global Ledger (XBRL GL) for detailed data and XBRL for Financial Reporting (XBRL FR) for end reporting. If structured properly, the very nature of this approach retains and makes available all the information present at each level, from the initial transaction (e.g., a customer order) to the related entries/documents (e.g., invoices and payments) to the various levels of aggregations (e.g., subledgers, ledgers, trial balance) to the end reports. And because it is based

on open, freely available standards and technologies, it also offers total ownership of the data and analytical processes (as opposed to relying on a specific vendor's format and architecture) as well as a faster, cheaper implementation.

Keeping a direct link between detailed data and the end reporting means:

◆ Always being able to reconcile data at each intermediate level and between different end reports;

◆ Not only having the ability to drill down from end report to detailed data and back, but also being able to "drill around" (tracking data as it flows to multiple end reports generated for different purposes); and

◆ Maintaining a seamless, unambiguous audit trail.

While data warehousing provides a repository of data that is controlled, internally consistent, and suitable as a starting point for generating analyses and end reports, it achieves its purposes by replacing the "real" detailed data with aggregation by predefined dimensions. Contextual information that isn't represented in those dimensions is lost in the process, resulting in a sort of lower boundary beyond which it becomes impossible to drill down to the original transactions. One of the most effective analogies to describe this comes from Eric E. Cohen, the creator of XBRL GL, who refers to it as the "roach motel": Data checks in to data warehouses, but it never checks out. Whenever a reconciliation process or some kind of analysis needs to use search criteria that aren't reflected in the predefined dimensions guiding the creation of the data warehouse, a manual process is the only solution. In this respect, data warehouses aren't very different

from spreadsheets. They are effective for their immediate purpose, but they generate a different version of the truth each time they are saved—with no possibility of reconciliation to the underlying data.

Some companies use data warehouses as their primary consolidation tool even though they were actually designed for analysis. XBRL GL-enabled systems can help meet the true need. Data warehouses are subject-oriented, but data integration needs to be subject-agnostic.

A Standards-Based Alternative

XBRL GL is a worldwide, holistic agreement on how to represent all kinds of business and accounting data in a consistent, application-independent way. It provides a standardized representation of the link between all levels of detailed data and end reports represented with XBRL FR or other XML schema-based reports. Sometimes it's necessary to link from XBRL GL to end reports that aren't XBRL but instead just based on XML schema. For example, certain regulatory filings, such as the U.S. Internal Revenue Service's tax forms, currently are based on XML schema but not XBRL.

XBRL GL is XML based, and as such it is the ideal payload for a Web services-based architecture capable of generating reports and analyses on demand. This helps avoid or minimize the physical duplication of data (in its original form and within the data warehouse), which is a great advantage in terms of streamlining the internal reporting process. Whenever data is duplicated, the need for additional reconciliations arises. Once duplicated, it immediately becomes obsolete unless a direct link with the underlying detail

can be maintained effectively.

XBRL also helps overcome another major obstacle of any architecture that involves moving data from one application to another: Each application has its own validation rules, different from each other, and sometimes in conflict. As data moves from one application to another, transformations are required to comply with the target application validation rules, and the context of the "source" validation rules is lost. With XBRL, it's possible to package validation rules with the data itself. The power of expressing validation rules in a standardized format, independent from source and target application, is obvious.

This doesn't mean that data warehouses are useless. They are suitable as extensions of end reporting toward some kind of dimensional detail—for example, when breaking down summarized information by some predefined relevant criteria, such as sales by region.

A direct link between detailed data and end reporting offers much more in terms of flexibility and agility in navigating the complex relationships between documents, entries, and different levels of internal and external reporting—something that data warehouses simply can't provide. The issues described here are brought about from trying to stretch the technology beyond its limits, and that ultimately leads to the data warehouse disconnect.

Without getting into a technical discussion, it's worth mentioning that even within XBRL there is space for both the approach based on predefined dimensions (data warehouses) and the approach based on the direct link between detailed data and reporting. Dimensional XBRL FR taxonomies are

similar to the data warehouse concept. Both are attempts to provide detail as support for end reporting through predefined dimensions rather than through being able to trace a direct link to the underlying detail. XBRL GL, on the other end, represents and maintains a full linkage between detailed and summarized data. Dimensional taxonomies aren't suitable for internal reporting uses—where reconciliation, drilling down, and tracking the audit trail are key. But they are, for instance, the architecture of choice within the current U.S. GAAP Project promoted by the Securities & Exchange Commission (SEC) and Financial Accounting Foundation (FAF) and developed by the XBRL US Jurisdiction. This is an architectural choice consistent with the scope and the space of that project, completely focused on external reporting. Just as data warehouses aren't suitable for consolidation and internal reporting purposes but have other applications, dimensional XBRL taxonomies are suitable for external reporting purposes but can't replace the direct link between end reporting and underlying transactions in internal reporting. XBRL GL ensures that it's still possible to maintain a direct link when end reports are represented with dimensional taxonomies through its SRCD (Summary Reporting Contextual Data) module, designed to allow a consistent, standardized representation of the link between XBRL GL and XBRL FR taxonomies, including dimensional taxonomies. The SRCD module of XBRL GL is now in Public Working Draft status (see www.xbrl.org/GLFiles), and I encourage everybody to review it and provide feedback to the XBRL GL Working Group.

Managing Internal Reporting

Solutions centered on data warehouses are perceived to be safe and reliable, but they are actually the source of many of the issues that management accountants face in their responsibility to design, implement, and manage internal reporting systems that support effective decision making. Innovative, standards-based approaches like the one described here are really the key to achieving the desired results while maximizing the long-term stability and reliability of the internal reporting environment. ■

Gianluca Garbellotto is an internationally known expert on both the business and technical aspects of XBRL and XBRL GL. He is the current chair of the XBRL GL Working Group and a member of the XBRL International Standards Board. Gianluca can be contacted at gg@iphix.net.

or what effect this "voice" is expected to have.

The second rule is to ascertain whether the goal will be achieved if the initiative is successful. It seems highly doubtful that shareholders petitioning the board of directors or even total divestment from any corporation doing business in Iran will have any effect on the attitude of the Iranian government toward the U.S. or its citizens or its terrorism activities. This is true for both private and public retirement and mutual funds.

The third rule is to concentrate shareowner activist efforts on the most important issues in which they can be successful. Good governance requires shareowners to actively monitor the actions of the corporations in which they directly or indirectly have an investment. Social issues of employee relations, community, diversity, suppliers, and the environment directly influence a corporation's daily operations. These are issues that management and the board of directors can control and that shareowners should be able to influence.

The benefit of hindsight shows that since 2005 the divestment statutes involving companies involved with Sudan don't appear to have solved or even influenced the problems in Darfur, and the government of Iran appears to be even more intransigent about terrorism and less subject to influence by outsiders.

Corporations can be expected to respond to only a limited number of issues. Shareholders' efforts should be concentrated on the most important issues: those that have a clear objective and where a successful initiative results in the goal being achieved.

BP Promises No Increased Pollution of Lake Michigan

In an August 23, 2007, press release, BP backed down from its plan to use its newly allowable Indiana permit to dump more ammonia and "sludge" solids into Lake Michigan (see "Is BP an Acronym for 'Big Polluter?'" September 2007). The company declined to make the pledge legally binding, however, further tainting its reputation for environmental concern. ■

Curtis C. Verschoor is the Ledger & Quill Research Professor, School of Accountancy and MIS, and Wicklander Research Fellow in the Institute for Business and Professional Ethics, both at DePaul University, Chicago. He is also a Research Scholar in the Center for Business Ethics at Bentley College, Waltham, Mass. His e-mail address is curtisverschoor@sbcglobal.net.