

TDWI

MONOGRAPH SERIES

SEPTEMBER 2009

Who Ensures Clean, Consistent Data? (Hint: It's Not Just the IT Department!)

By Wayne Eckerson
Director, TDWI Research
The Data Warehousing Institute

SPONSORED BY



About TDWI Research

TDWI Research provides research and advisory services to business intelligence and data warehousing (BI/DW) professionals worldwide. TDWI Research focuses exclusively on BI/DW issues and teams up with industry practitioners and TDWI faculty members to deliver both a broad and deep understanding of the business and technical issues surrounding the deployment of BI/DW solutions. TDWI Research offers reports, commentary, and inquiry services via TDWI's worldwide Membership program and provides custom research, benchmarking, and strategic planning services to both user and vendor organizations.

About TDWI

The Data Warehousing Institute (TDWI), a division of 1105 Media, Inc., provides in-depth, high-quality education, training, research, and certification for business intelligence (BI) and data warehousing professionals worldwide. TDWI can help your BI team stay abreast of new and emerging trends and techniques and gain the skills they need to deliver effective BI and data warehouse solutions. Through our Membership program and regional chapters, TDWI can also help you and your team establish a network of peers in the industry to whom they can turn for assistance and advice in career development and to solve thorny technical and organizational problems. www.tdwi.org

About SAS

SAS is the leader in [business analytics](#) software and services, and the largest independent vendor in the business intelligence market. Through innovative solutions delivered within an integrated framework, SAS offers unparalleled data integration and data quality, is the market leader in predictive analytics and provides sophisticated reporting. SAS helps customers at more than 45,000 sites improve performance and deliver value by making better decisions faster. Since 1976 SAS has been giving customers around the world The Power to Know®.

© 2009 by TDWI (The Data Warehousing Institute™), a division of 1105 Media, Inc. All rights reserved. Reproductions in whole or in part are prohibited except by written permission. E-mail requests or feedback to info@tdwi.org.

Product and company names mentioned herein may be trademarks and/or registered trademarks of their respective companies.

Executive Summary

Many of the problems associated with business software applications are caused by problems with data, more specifically data defects and data inconsistency. Business executives and managers need to understand the impact that poor data can have on their business initiatives and programs. Rather than blaming the information technology (IT) department for these problems, the business needs to recognize that data quality and consistency are a joint responsibility.

Top executives need to view data as a valuable corporate asset, assign senior managers to oversee data governance initiatives, make subject matter experts available to assess and resolve data issues, and partner with the IT department to implement its policies and standards regarding data. In short, the business establishes policies and standards for enterprise data and the IT department implements them.

A strong business-IT alliance ensures that an organization maximizes the value of its information assets and avoids problems due to dirty or inconsistent information.

The Business Imperative

In 2002, The Data Warehousing Institute (TDWI) estimated that inaccurate customer data costs U.S. businesses a staggering \$611 billion a year in postage, printing, and staff overhead.¹ Frighteningly, the real cost of bad data is higher. Data problems can alienate customers, create revenue and cost leaks, undermine process efficiency, delay expensive projects, and expose an organization to compliance risks. In short, bad data can make it hard for the business to achieve its financial and strategic goals.

Consider the following examples:

- A telecommunications firm lost \$8 million a month because data entry errors incorrectly coded accounts, preventing bills from being sent out.
- A global chemical company discovered it was losing millions of dollars in volume discounts in procuring supplies because it could not correctly identify and reconcile suppliers on a global basis.
- An insurance company lost hundreds of thousands of dollars annually in mailing costs (postage, returns, collateral, and staff to process returns) due to duplicate customer records.
- An information services firm lost \$500,000 annually and alienated customers because it repeatedly recalled reports sent to subscribers due to inaccurate data.
- A large bank discovered that 62% of its home equity loans were being calculated incorrectly, with the principal getting larger each month.
- A health insurance company in the Midwest delayed a decision support system for two years because the quality of its data was “suspect.”

Despite the risks, many business executives don't understand the high costs of bad data. Or if they are aware of the problems, they don't know what steps to take to resolve them. IT departments feel helpless to resolve the issue without strong executive sponsorship and funding. Compounding the problem, there is a communications barrier between the two groups. Business folks often feel that IT people speak a different language (“IT speak” versus “business speak”), and IT developers often find that business executives and experts won't devote the time needed to help them understand the business content of the data.

The reality is that the business runs on data.

The reality is that the business runs on data; it's like fuel for the corporate engine. Without good data, a company can't possibly understand its customers, suppliers, or competitors—or its own people, processes, and performance. Therefore, it's imperative that business and IT find common ground and work together to ensure high-quality data. For example, the business must establish

¹ TDWI's estimate was based on cost savings cited by survey respondents and others who have cleaned up name and address data, combined with Dunn & Bradstreet counts of U.S. businesses by number of employees.

standards for the definition of common business elements, such as “customer” or “sale,” and common calculations including currency conversions and employee costs. Once these terms and rules are defined, the IT department then encodes them in software and checks all data flows to ensure adherence to the standards.

Types of Data Problems

There are many types of data problems that afflict businesses, but they can generally be divided into two camps: defective data and inconsistent data. Unfortunately, most companies don’t take the necessary steps to address these issues until they reach crisis proportion and threaten to jeopardize the business.

Defective Data

There are many types of data defects.

Types of Data Defects. There are many types of data defects. Data that is inaccurate, incomplete, unavailable, or obsolete can be considered defective. For example, most people know that manually entered data—whether entered by hand by internal administrators or by customers filling out forms on the Web—is often full of spelling errors, typos, and perhaps falsified data (for example, “Mickey Mouse” for a name.) Also, a person may neglect to fill in all of the fields or enter correct data in the wrong fields. Data defects also appear when a system is moved from one platform to another, an old application is replaced with a new one, or data is moved from one application to another on a scheduled and automated basis.

Fixing Data Defects. The problem with defective data is that it’s difficult and expensive to root out once it’s entered into the system. The best way to remedy defective data is to prevent it from happening in the first place. This requires the business to invest in systems that validate and fix data at the source—at the point when it’s entered into the system or moved between systems via an application interface. To fix defects already in the system, the business must invest in data profiling and cleansing tools to scrub data sets before they are loaded into data warehouses or master data management systems. Data stewards from the business define the rules for fixing defects and standardized data elements that are implemented by their IT counterparts.

Inconsistent Data

External data has a short shelf life.

External Data. A more pernicious problem than defective data is inconsistent data. Data has a short shelf life. Like a radioactive substance, it seems to decompose quickly into distinct elements that can’t easily be reconstructed into the original. For example, customer data degrades over time because people marry, divorce, die, move, or change their names. Experts say that 20% of customer databases contain duplicate or inconsistent data.

Internal Data. Internally, companies excel at fragmenting their own data, largely because of the way they are organized and also because they lack centralized data

management. Most CEOs have experienced the “dueling spreadsheet” phenomenon in which senior executives empowered with spreadsheets created by business analysts spend an entire meeting arguing about whose data is correct rather than making decisions. Each analyst, department, or division sees the world through its own lens and thus defines and models data differently. In such an environment, it’s common for the heads of finance, marketing, sales, and engineering to disagree about the definition of something as basic as “customer” or “sale.”

Continuous Vigilance. Managing the consistency of external data—such as customer data—requires continuous vigilance by the IT department, which needs to be equipped with tools for verifying, standardizing, matching, and householding data to avoid duplicates. And managing internally inconsistent data is politically challenging; it usually requires the CEO to force executives to come to a consensus on the meaning of key terms and then appoint business managers to serve as stewards of those definitions and the data associated with them. In addition, the CEO needs to fund an enterprise data warehouse (EDW) and master data management (MDM) systems that enforce the definitions and rules so the organization maintains a “single version of truth.”

Solving Data Problems

We just described the common data problems that undermine the ability of companies to conduct business. Let’s now look at real cases and discover how business and IT worked together to deliver clean, consistent data that benefited their organization in numerous ways.

Reducing Defects Via Stewardship

Embarrassed by the snafu, the CEO established a data stewardship and quality program.

Dear Mexican. A regional bank in the Southwest found out the impact of defective data the hard way. The CEO received an angry phone call from the head of the Mexican consulate—a major customer of the bank—because he had received a direct mail letter from the bank addressed “Dear Mexican.” Highly embarrassed by this snafu, the CEO established a data stewardship program to fix data quality problems that were adversely affecting client relationships and jeopardizing a customer relationship management (CRM) system that the bank had recently invested significant time and money to implement.

Objectives. The CEO then laid out the objectives for the initiative in a company newsletter:

Our key objective ... is to materially improve the profitability of the company by 1) ensuring decisions are based upon fresh and accurate information, and 2) reducing expenses associated with the inefficiencies incurred from rework, complaint handling, elongated processing time, unplanned service, and redundant efforts.

According to one executive, the company’s goal was to achieve “zero defects” and continuously measure its progress toward that goal, along with the costs of

variance. The bank also set forth expectations to employees: “The journey to quality information does not happen overnight. The process to correct and prevent recurrences can be lengthy.” The company’s BI department created a dashboard to track defects and progress toward achieving its goals.

Oversight Tasks. The next thing the CEO did was to invite senior managers from all lines of business to serve as chief data stewards in their areas. The corporate stewardship committee provided direction to all data quality teams scattered throughout the company. Specifically, the committee:

- Educated workers on the importance of data quality to the company
- Communicated data quality improvements to all employees
- Defined mandatory data elements that needed to be measured
- Reviewed metrics for measuring the quality of these elements
- Defined methods of reporting on data quality levels
- Set precedence for establishing service-level agreements for data quality
- Established the “owners” and “custodians” of each major subject area and data store
- Resolved cross-divisional enterprise data quality issues
- Ensured that code sets would be updated regularly

In essence, the corporate stewardship committee became a clearinghouse for ideas and information pertaining to data quality. The stewardship committee also worked closely with the IT department, which helped significantly in the effort. The IT department built a dashboard to track the progress of the data quality initiative, and implemented the business rules for capturing and fixing data defects and standardizing the semantics and formats of key data elements, such as customer, product, and sales. By working closely together, the business and IT department helped the regional bank make significant headway in achieving its data quality goals within 18 months.

The stewardship committee worked closely with the IT department.

Delivering Consistent Data Through BI

Major U.S. Railroad. While fixing data defects requires steady, diligent effort, fixing inconsistent data can be politically charged. The operations department of a major U.S. railroad company recognized that it needed business intelligence to ensure a single version of truth and streamline operations meetings.

The railroad’s operation department holds a daily meeting at 8 a.m., attended by the chief operations officer, three regional vice presidents, and divisional managers. Several years ago, the meetings revolved around a printed 120-page daily performance report consisting of tables of data (and no charts) that examined 45 metrics across all products, organizations, and geographies. Prior to each meeting, operations analysts for each executive and manager crunched the numbers in the report and from elsewhere to answer questions about the daily performance report.

“The analysts summarized the data differently, and often the details did not add up to the total.”

Different Views. “The analysts from different departments and regions summarized the data differently, and often the details did not add up to the total on the daily performance report,” said the company’s director of BI. He added that the managers spent more time in the meeting debating the validity of their numbers than determining the best actions to take to fix problems. “Finally, the COO said, ‘There has to be a better way,’ and he chartered us to turn the printed performance report into an interactive, corporate dashboard that would let each department view the data from its local perspective while preserving an overall corporate view,” said the BI director.

BI Initiative. The BI team spent seven months and \$500,000—mostly in labor—to build the dashboard, which has been a huge success. The dashboard displays metrics using heat map visualizations, geographic mapping tools, and OLAP client software, which all run against an OLAP database that extracts data from the corporate data warehouse. The application has been a tremendous success. It has streamlined operations meetings, helping the team pinpoint problems and understand their impact across all departments and regions. The COO has continued to fund the expansion of the dashboard, and other departments at the company are now requesting similar capabilities.

Managing Expectations. Interestingly, the BI team encountered a serious data quality problem when constructing the dashboard. It discovered that a key metric in the daily performance report was calculated incorrectly and delivered erroneous data. The BI team decided not to fix the data defect prior to the launch of the dashboard, fearing that a disagreement over the veracity of the data behind a key metric would jeopardize or delay the entire project. So, it consciously delivered erroneous data to manage business user expectations, and only later, once the project gained acceptance, did it alert the business owners of the faulty data and made the requisite changes.

Health Payer. Another example of an organization whose business executives understood the importance of delivering consistent information is a major health insurance provider in the Midwest. “Every health care provider says they are in the business of selling insurance, but really, we are in the information business,” says a vice president of enterprise information management.

Unfortunately, the company’s main data—claims, member, and provider information—was embedded in every application of the company. This not only made it impossible to keep the data in sync, but it also created a Web of one-to-one linkages between applications that were expensive to maintain and caused performance problems (that is, when one system was down or running slow, it degraded performance in all the other systems.)

The initiative to address these data issues started at the very top.

The initiative to address these data issues started at the very top. The CEO laid out new strategy and performance goals, and part of the strategy was to “leverage knowledge capital.” Putting weight behind the proclamations, the CEO funded the development of an enterprise asset strategy with a mission and road map that called for the creation of a data governance and stewardship program to manage data consistency. The strategy also called for a master data management (MDM)

initiative to standardize and centralize claims, provider, and member information. To oversee the governance program and MDM initiative, the CEO created a new executive position with the title of senior vice president for strategy and data management. This new executive, in turn, hired a team to implement the MDM initiative.

The MDM initiative created a single repository of claims, member, and provider data for all applications to access. This ensures that the company maintains a single standardized version of these core data elements. It also relieves the claims, member, and provider systems of dependencies with other systems and offloads queries that could hamper performance. To create standardized master data, the team created a logical model with input from the business data stewards in each department. “They help us define what every field means so every data element has a single definition and format in the repository,” said the EIM manager. “We don’t define anything; we just let the stewards in each business unit fight it out.”

Bridging Business and IT

Business must take an active role to improve the quality of data.

It’s clear from the case studies that business must take an active role in both the creation of data quality standards and in working with IT to improve the quality of data, which is a critical asset for many companies today. Typically, after a business initiative or application has been implemented, companies don’t pay much attention to the quality and consistency of their data until a crisis or opportunity occurs. The regional bank didn’t recognize its data problems until a major customer lodged an angry complaint; the railroad company didn’t build an operations dashboard until it could no longer conduct its daily meetings effectively; and the Midwestern health insurer didn’t shore up its data foundation until process inefficiencies began undermining its ability to achieve strategic objectives.

Starts at the Top. In all cases, data strategies started at the top, with a chief executive officer or senior executive. What’s key is that they didn’t throw the problem over the wall to IT; instead, they appointed senior business managers to oversee the design and implementation of a data governance strategy and held them accountable for achieving well-defined outcomes for improving data quality and consistency. Those managers in turn appointed business users to serve as stewards of critical data elements in their domain, working with other stewards to harmonize data definitions and formats, manage exceptions, and validate the quality of relevant data in reports and applications. They also provided funds to the IT department to implement BI and MDM initiatives.

Grassroots Alternatives. However, if executive support is not forthcoming for a data quality or governance initiative, don’t despair. Business and IT can still work together to begin building the foundation for clean, consistent data at the grassroots level and work their way up to the top. This means that business subject matter experts and IT developers need to work together to embed data quality within projects. Before tackling a new project, reach out across

departmental lines and partner with the relevant representatives to align data definitions and models to ensure seamless extension of the solution to the interested department.

Bake data governance processes into project applications and reviews.

Another grassroots opportunity is to work with the project management office or a BI or data integration competency center to bake data governance processes into project applications and reviews. Also, since most data initiatives involve fixing source system processes or data entry validation, establish a data quality center within IT to raise consciousness about the impact of poor data quality on the business and identify ways to ameliorate the problems.

Data is the foundation of most modern businesses. Senior executives must take responsibility for managing data as a corporate asset and work together with IT to spearhead initiatives to deliver high-quality, consistent data. In doing so, future initiatives will be more successful by avoiding the costs, delays, and embarrassment frequently associated with bad data.