With data management, it's garbage in, garbage out. Do your processes—and consequently your business decisions—stink?

Aawww, RUBBISH!

By Rick Whiting

HOME IN THE SMALL TOWN of Valparaiso, Ind., valued at $121,900 somehow wound up recorded in Porter County's computer system as being worth a whopping $400 million. Naturally, the figure ended up on documents used to calculate tax rates. By the time the blunder was uncovered in February, the damage was done: Valparaiso, its school district, and government agencies were forced to slash their budgets by $3.1 million when they found they wouldn't be getting the tax dollars after all.

It's a nightmare scenario—and one like it could be yours. Bad data remains a major cause of botched marketing campaigns, failed CRM and data warehouse projects, angry customers, and lunkhead decisions. Despite all we know about the importance of data scrubbing and quality management, many companies are still using data that's redundant, incomplete, conflicting, outdated, and just plain wrong.

Bad data isn't a new problem, but urgency in dealing with it is at an all-time high. Customers are voicing anger at the mistargeted marketing pitches and poor service that result from off-the-mark data, and they're taking their business elsewhere. Companies are investing billions of dollars in CRM applications and data integration projects to gain a better view of their customers—only to discover that conflicting data makes them blind. "Our marketing effectiveness leads to our sales effectiveness, which leads to our service effectiveness. Data quality is key to the success of that," says Chuck Scoggins, VP of customer solutions at Hilton Hotels. "If you don't have quality data, that whole chain breaks down."

Illustration by Jay Montgomery

IN DEPTH / DATA QUALITY

Managers and employees increasingly base decisions on insights gleaned from performance management applications and dashboards. But business intelligence tools are only as good as the data that goes into them; faulty data leads to ill-informed decisions. The ramifications range from ticked-off customers to misled investors to testy regulators. Executives can face jail time under the Sarbanes-Oxley Act if they don't have financial data in order. Bad data can even increase the cost and time involved in completing mergers by making it more difficult to integrate operations and combine customer lists.

The problem is getting harder to manage as the amount of data generated and maintained by many businesses doubles every 12 to 18 months. And as more businesses share information with outside partners and customers, more bad data is being exposed to others. Lax quality is familiar to anyone with a mail box: Consum- ers get credit card pitches from issuers with which they already have cards, mailings from charities in triplicate with slightly different name spellings, and warranty extension offers from auto dealers for cars they no longer own.

Occasional inconvenience for consumers aside, low-quality data is foremost a problem for the company holding it. Bad data can be an embarrassment—companies are loath to talk openly about internal data disasters. Businesses may be legally bound to share information about security breaches that result in consumers' personal information being compromised, but that's not the case with bad data. As a result, tales of mishaps are hard to come by, even as the problem persists.

The biggest obstacle to fixing the mess is that business managers view data quality as a technical problem, when business processes are really what's broken. IT has little control over the sales rep who gets a customer address wrong on an order or the manufacturing manager who enters an incorrect part number in an inventory database. A Gartner survey of 600 executives in November found that 49% think the IT department is responsible for their organizations' data quality; much smaller numbers say responsibility lies with top execs, data quality

QUALITY EFFORTS REQUIRE TEETH

AVAYA MAINTAINS MORE THAN 100 terabytes of customer, vendor, service, financial, and pricing data. To ensure that the stockpile complies with internal data standards, the company, which provides telecom equipment and services, last year established its Data Quality Center of Excellence.

The center's two dozen employees are responsible for implementing data quality management practices, such as avoiding the creation of duplicate records. Championed by Guy Lardieri, VP of strategic initiatives and business architecture, the center was created in April 2005 as a spin-off from a project to replace an aging system with enterprise applications from SAP and Siebel Systems.

The legacy system was hamstrung by defective data that drove up expenses and cut into revenue. In some cases, Avaya serviced customers' telecom equipment but didn't bill them for gear that was erroneously left out of service agreements. Other customers who paid only for standard service were getting premium service because of database errors, says Rich Trapp, Avaya's global data quality director.

Avaya's Data Quality Center of Excellence provides the tools for improving and maintaining data quality, including Business Objects' IQ Insight profiling tool for identifying bad data. The Data Quality Executive Council, made up of top company executives, "provides the teeth for the data quality efforts," Trapp says. But it's the business units that have ultimate responsibility for data quality. **—RICK WHITING**

Bad data cost Avaya good money, Trapp says.

teams, line-of-business managers, and others.

"Business has to accept the fact that it has primary responsibility for data quality. Data is a business asset," says Nigel Turner, who as project lead manager for data quality programs at BT Group (formerly British Telecom) in the late '90s helped get that company's data cleanup efforts off the ground.

Gartner estimates that more than 25% of critical data within large businesses is somehow inaccurate or incomplete. And that imprecise data is wreaking havoc. Fifty-three percent of the 750 IT professionals and business executives surveyed by the Data Warehousing Institute late last year said their companies had experienced problems and suffered losses or increased costs because of poor-quality data, up from 44% in a similar survey in 2001.

While IT managers may not own the processes that spew bad data, they can make the business case to change those processes to improve data quality. Moreover, they can provide the technology to support those improved processes and, since no process is perfect, operate the tools needed to automate the downstream steps of identifying and correcting bad data.

DATA QUALITY CHAMPIONS

Though BT began adopting data quality practices 20 years ago, its real effort began in 1997 as it struggled with customer billing errors and poor product inventory information. The company's efforts to improve interactions with suppliers and customers using EDI and self-service applications were also being hindered by bad data.

Turner, then in BT's corporate strategy division, recognized that the telecommunications company was spending a great deal of effort correcting data. Rather than create a top-down, companywide program, Turner targeted line-of-business operations and identified a data quality "champion" in each to lead an information management forum. The groups targeted specific projects with demonstrable returns on investment, such as improving names and addresses in marketing data to reduce the number of letters sent to the wrong people and improving private-line inventory record keeping to increase the number of disconnected circuits re-

Quality's Upside
What benefits has your company derived from high-quality data?

Benefit	%
Greater confidence in analytic systems	76%
Less time spent reconciling data	70%
Single version of the truth	69%
Increased customer satisfaction	57%
Reduced costs	56%
Increased revenue	30%
Other	5%

% of respondents

Note: Multiple responses allowed.

Data: Data Warehouse Institute 2005 survey of 750 data management professionals

turned to stock for reuse.

"We had to prove to BT that these things were worth doing," Turner says. "Data quality isn't very sexy." The original budget for the data quality efforts was a measly $30,000. As the project expanded, Turner's group developed a data quality methodology incorporating best practices gleaned from inside the company and from outside experts, and centralized data quality management. Recognizing that errors will creep into databases despite its best efforts, BT uses data profiling and cleansing tools from Trillium to identify and remove errant data.

The efforts have paid off: BT has realized as much as $800 million in aggregate savings by improving inventory management, boosting productivity through improved automated interactions with suppliers and customers, and reducing revenue leakage through more accurate billing. BT has parlayed its data quality know-how into a consulting business headed by Turner.

Still, data quality problems are legion and seem to exist to some degree at all manner of companies that manage large quantities of information. Darren Cunningham, product marketing director at Business Objects, shares the story of a consumer technology manufacturer that routinely sent half of its catalogs to the wrong addresses until a manager pointed out the high number of catalog returns and customer complaints. Taking steps to correct the problem saved the company $12 million a year, Cunningham says.

Data quality initiatives can be part of broader data governance programs. Data governance, a relatively new concept, applies best practices to how information is managed, secured, and used across an organization. It requires establishing a formal set of business processes and policies to ensure that data is handled in a prescribed fashion. Data governance includes standard definitions for data elements to be used throughout a company—just what a "lost customer" is, for example—and metrics for measuring data quality, says Terry Haas, director of the enterprise data management practice at PricewaterhouseCoopers. Data governance also defines the data management roles and responsibilities of managers and employees and
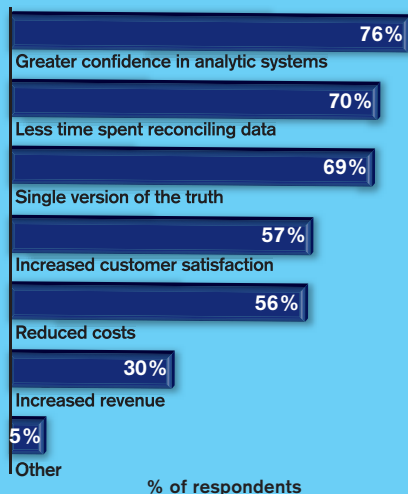
▶ IN DEPTH / DATA QUALITY

limits the ability to change data to designated "data stewards."

There's no standard way of measuring data quality. Bank of America and Cintas use Six Sigma as a yardstick. (Six Sigma is a methodology for measuring and removing defects from everything from data to manufactured products.) Hilton Hotels uses the probability of correctness indicator, or PCI, which assigns data a rank of one through nine based on its trustworthiness. Hilton rates 95% of its customer data at the high end, in the one through four categories. But, reflecting an emphasis on measuring data quality projects by their ROI, BT's Turner says the one metric that matters is money.

DUPLICATION DANGERS

Bank of America has long collected account data in a centralized data warehouse for a variety of marketing and cross-selling applications. The bank's data quality efforts began in earnest in 2002 to comply with the anti-money-laundering provisions in the USA Patriot Act. Data on new accounts is collected in

Database Debacles

Has your company suffered losses, problems, or costs because of poor-quality data?

53% Yes
36% Haven't studied
11% No
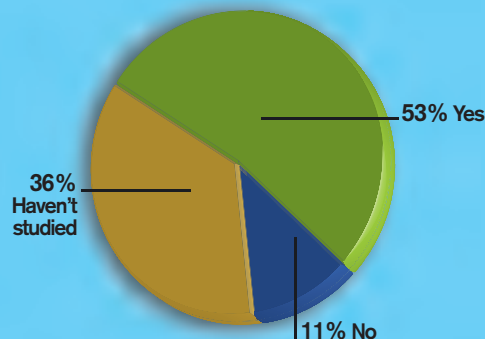
Data: Data Warehouse Institute 2005 survey of 750 data management professionals

the multiterabyte warehouse from several lines of business, so Bank of America established common practices for capturing, integrating, and managing it, says Donald Carlson, who heads up the bank's anti-money-laundering program and has become its de facto data quality manager.

The bank designated data stewards in business units and the IT department, and some with companywide responsibility. Data quality managers meet monthly to resolve problems. Bank of America uses commercial and custom-built data profiling and matching tools to examine and, when necessary, correct data sent to the warehouse. Today, in addition to regulatory compliance, the bank's data quality efforts are driven by its risk management practices, the need to manage customer data from multiple channels, and cross-selling efforts.

Integrating data from multiple business operations has also been a challenge at Cintas, which created new divisions as it expanded beyond its core employee uniform business into areas such as providing businesses with cleaning supplies and document storage and

→ ACCURACY GETS BAKED IN

THERE'S A SOFTWARE tool for every nitty-gritty data quality management task, and big tech vendors are acquiring specialists to piece together a complete set of products.

Business Objects, IBM, and Informatica have snapped up software companies with tools they think can help inject data quality management into broader IT and business processes. Business Objects last month completed a $69 million buyout of Firstlogic, a supplier of data quality software and services. The Firstlogic technology has become a key element of the vendor's enterprise information

management line of products and services, unveiled last week and designed to provide a foundation for its business intelligence and performance management applications.

IBM last year acquired Ascential Software, including its ProfileStage data-profiling software, and a toolset for data matching and de-duping called QualityStage. Now the software is part of IBM's WebSphere Information Integration system. In January, Informatica added data quality tools to its PowerCenter data integration software when it bought Similarity Systems, a developer of data profiling, cleansing, matching, and stan-

dardization software. Two years ago, Pitney Bowes bought Group1 Software, another data quality tool vendor, and began offering its technology for a range of address correction, customer data integration, and geospatial applications.

The risk of integrating data quality tools into other products is that it reduces visibility of the problem. But the trend helps put data quality where it belongs—in the middle of IT processes and the business processes they support. Says Gartner's Ted Friedman, "It's made the synergy between data quality and data integration more obvious."
—RICK WHITING

shredding services. That has resulted in customer data silos throughout the company, database marketing manager Becki Wessel says.

To help with cross-selling, data from all divisions is collected in a data warehouse, but the information is sometimes duplicated with slight variations. Some customers are listed in multiple databases but with enough variation in name or address to be identified as different people. Those discrepancies have sometimes led to existing customers being identified as new prospects—an embarrassing situation when a sales rep shows up. An added danger is that sales reps could begin to distrust leads provided by marketing, Wessel says. Or two customers could be close enough in spelling to be tagged as the same customer, costing the company a sales opportunity.

As part of a project to overhaul its data warehouse, Cintas has been installing quality management software from Dataflux that will identify duplicate customer records and standardize customer data collected monthly from each division's database. The system is expected to be fully functional by next month, but a pilot project already has improved the company's ability to match customer names.

While Cintas is integrating customer data on a monthly batch basis, other companies do so in real or near-real time, which makes data quality even more difficult. More companies also are adding third-party data that may be erroneous or inconsistent. Bank of America's Carlson notes that the globalization of business—and data sources—further complicates the problem.

WHEN GOOD DATA IS THE BOTTOM LINE

DATA QUALITY IS IMPORTANT to any business. But when data is your business, as it is for database marketers, it has to be squeaky clean.

Information Resources Inc. has 170 terabytes of data, the majority of it point-of-sale information collected from 40,000 grocery, drug, and convenience stores—600 million records every week—on what packaged goods and health care products consumers are buying. That's supplemented with data provided by 100,000 households and information about the promotional activities of consumer packaged goods manufacturers and retailers. The data is sold to consumer packaged goods manufacturers and retailers, who use it to calculate market share, plan marketing campaigns, and carry out other analysis.

IRI corrects and standardizes data as it comes in from retailers in 1,850 formats, CIO Marshall Gibbs says. The company uses IBM's DataStage for quality control and management and Ilog's JRules rules engine to apply the necessary business rules to data. IRI uses its own algorithms for detecting any anomalies.

The key to IRI's success is using Six Sigma methods to analyze data collection and management processes before implementing data quality automation software. "Automating a process works much better after you've done the process reengineering to improve it," Gibbs says, adding that the amount of bad data that gets through to customers is negligible.

Acxiom, which augments data provided by its business customers with terabytes of consumer data that it collects and maintains, monitors the quality of data coming from clients using a scorecard system and metrics developed internally in the late 1990s, says Traci Campbell, a data quality consultant at Acxiom. Incoming data is scrutinized for accuracy and completeness, proper categorization, and redundancies before any corrective action is taken.

Problems still crop up, such as when a client prepares consumer data for a project, then tries to use that data for another project for which it's missing key information or isn't properly formatted. And it's a challenge to keep consumer data fresh, reflecting up-to-date address changes, marriages, divorces, and other life-changing events, says Bob Gillen, address quality services product manager. "It's not something you fix once."
                                                                    —RICK WHITING

BLIND DATA MATCHING

In the hotel business, the data challenge is exacerbated by the fact that customers don't need to use their real names to make a reservation. The Hilton hotel chain's 4-terabyte data warehouse stores the names of 22 million customers who have stayed at a Hilton multiple times over two years and 60 million identified as infrequent guests. Data about mem-bers of the Hilton Honors frequent guest program (about 20% of all Hilton guests) is included in the 22 million, and their data is presumed to be accurate, Scoggins says.

But it's difficult to sort out which category to put all the other guests in when they make a reservation. Hilton uses a combination of custom-built tools and software from Group 1 (owned by Pitney Bowes) to match a guest's name and address with information already in the database. That includes a Soundex algorithm that matches names based on phonetic pronunciation

rather than spelling. Only 40% of all customers are matched with an existing pro-file, and new profiles are created for the rest, Scoggins says.

Master data management projects tied to CRM, ERP, and supply chain management systems are some of the biggest drivers of data quality programs, AMR Research analyst Bill Swanton says. Master data management involves using a centrally managed database of customer names, product numbers, and other critical data. "Typically, we see people getting the data quality religion because they implement a big, expensive IT project and it doesn't work," Swanton says.

At BMW Group Canada, customer data is generated by retail and call center operations, company-sponsored events, and direct mail and Internet marketing campaigns. Since June, the company has been centralizing all that customer data in a Siebel CRM system and using data matching software from Trillium to eliminate duplicate customer records, standardize names and addresses, and fill out incomplete records. Before the system was installed, dealers, BMW financial services, and other operations had their own customer databases, and customers complained that they

had no central point of contact, marketing services manager Kelly Lam says. The company is also saving on mailing costs by complying with address format standards set by Canada Post, he says.

But the ultimate goal of data quality improvement is to catch errors at the point of entry or, even better, prevent errors from occurring at all, says Philip Russom, senior manager of research and services at the Data Warehousing Institute.

Some companies already are thinking along those lines. Cintas is considering using the data matching capabilities in the Dataflux system to correct data on the fly as divisional employees enter it into the system, rather than when it enters the data warehouse. Accurate data starts at the beginning—and the work required to keep it clean never ends. Says Wessel: "As long as you're fixing things on the back end, you're not correcting the problem."

Write to Rick Whiting at rwhiting@cmp.com.
Visit Database Pipeline: informationweek.database
pipeline.com