

Data Quality and Why We Should Care

Clean data don't just make actuaries' jobs easier, they can make the difference between profit and disaster.

CARING ABOUT DATA QUALITY is the key to safeguarding and improving it. But can we, as the expression goes, “recognize it when we see it”? Considerable analysis and much experience make it clear that the answer is no. Discovering whether data are of acceptable quality is a measurement task, and not a very easy one. But it's even more important in this information age, if information isn't to become misinformation.

Data are of high quality if they're “fit for use” in their intended operational, decision-making, and other roles. In many settings, especially for intermediate products, it's also convenient to define quality as “conformance to standards.” These two criteria link the role of the employee doing the work (conformance to standards) to the client receiving the product (fitness for use). When used together, these two criteria can yield efficient systems that achieve the desired accuracy level or other specified quality attributes.

Unfortunately, the data of many organizations don't meet either of these criteria. As the cost of computers and computer storage has plunged over the past 50 or 60 years, the number of databases has skyrocketed. With the wide availability of sophisticated statistical software and many well-trained data analysts, there is a keen desire to analyze such databases in depth. Unfortunately, after they begin their efforts, many data analysts realize

that their data are too messy to analyze without major data cleansing.

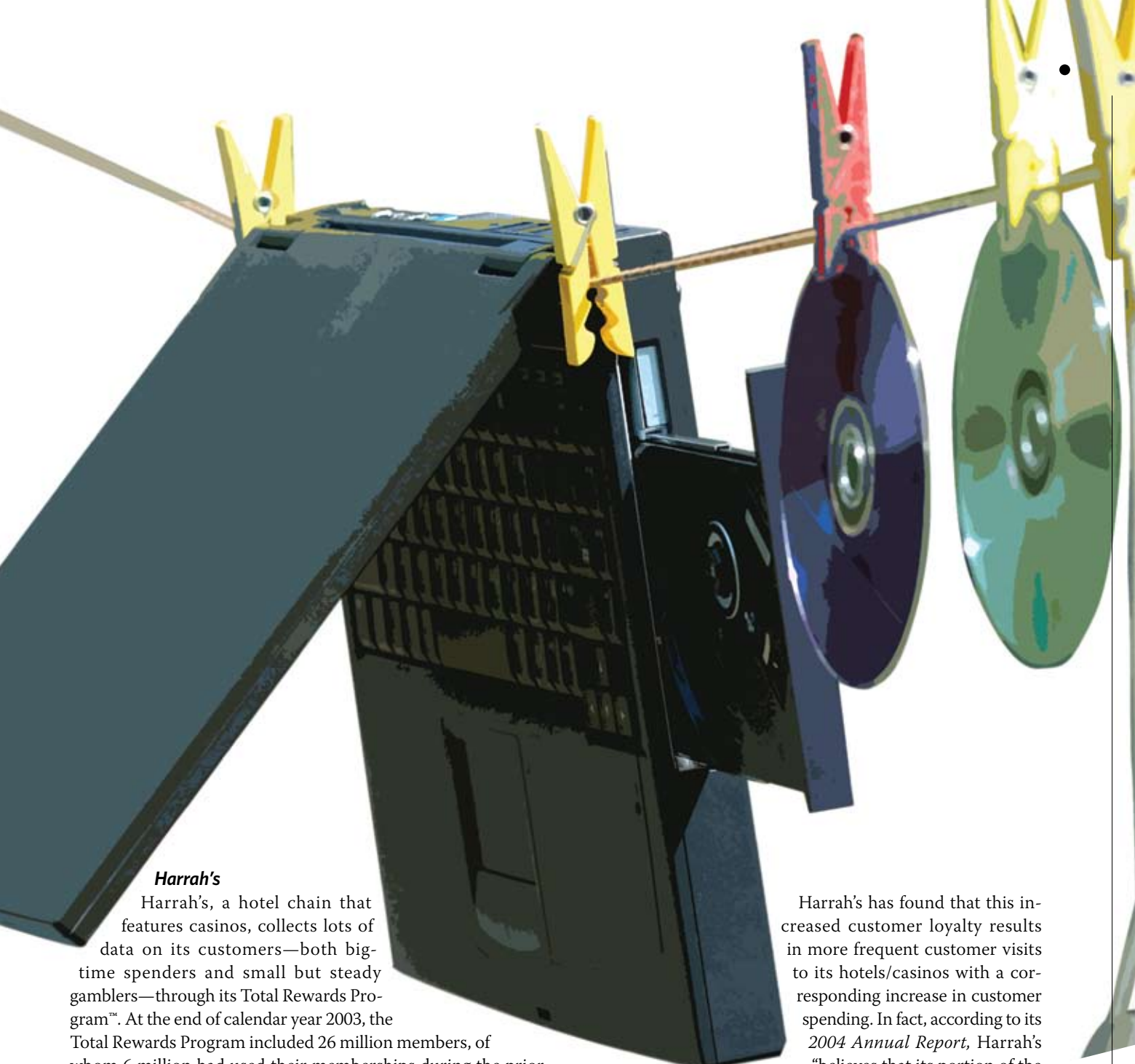
Data quality is important to business and government for a number of obvious reasons. First, a reputation for world-class quality is profitable, a “business maker.” As the examples below show, high-quality data can be a major business asset, a unique source of competitive advantage.

By the same token, poor-quality data can reduce customer satisfaction. Poor-quality data can lower employee job satisfaction too, leading to excessive turnover and the resulting loss of key process knowledge. Poor-quality data can also breed organizational mistrust and make it hard to mount efforts that lead to needed improvements.

Further, poor-quality data can distort key corporate financial data, making the financial condition of a business impossible to determine. The prominence of data-quality issues in corporate governance has become even greater with enactment of the Sarbanes-Oxley legislation, which holds senior corporate management responsible for the quality of a company's data.

Data Quality as a Competitive Advantage

In the examples that follow, we show how timeliness, accuracy, comparability, and completeness of data can yield competitive advantage.



Harrah's

Harrah's, a hotel chain that features casinos, collects lots of data on its customers—both big-time spenders and small but steady gamblers—through its Total Rewards Program™. At the end of calendar year 2003, the Total Rewards Program included 26 million members, of whom 6 million had used their memberships during the prior 12 months.

The database for this program is based on an integrated, nationwide computer system that permits real-time communication among all of Harrah's properties. Harrah's uses these data to learn as much as it can about its customers in order to give its hotel/casino guests customized treatment. This enables Harrah's to know their gambling, eating, and spending preferences. Hence, Harrah's can tailor its services to its customers by giving customized complimentary services such as free dinners, hotel rooms, show tickets, and spa services.

While the prevailing wisdom in the hotel business is that the attractiveness of a property drives business, Harrah's further stimulates demand by knowing its customers. This shows that Harrah's is listening to its customers and helps Harrah's to build customer loyalty.

Harrah's has found that this increased customer loyalty results in more frequent customer visits to its hotels/casinos with a corresponding increase in customer spending. In fact, according to its *2004 Annual Report*, Harrah's "believes that its portion of the customer gaming budget has climbed from 36 percent in 1998 to more than 43 percent" in 2002.

Wal-Mart

According to Wal-Mart's *2005 Annual Report*, Wal-Mart employs more than 75,000 people in logistics and in its information systems division. These employees enable Wal-Mart to successfully implement a "retailing strategy that strives to have what the customer wants, when the customer wants it."

"With the Data Warehouse storage capacity of over 570 terabytes—larger than all of the fixed pages on the Internet—we [Wal-Mart] have [put] a remarkable level of real-time visibility planning into our merchandise planning. So much so that when Hurricane Ivan was heading toward the Florida panhandle, we knew that there would be a rise in demand for Kellogg's Straw-

THE RISK OF MASSIVE ID FRAUD

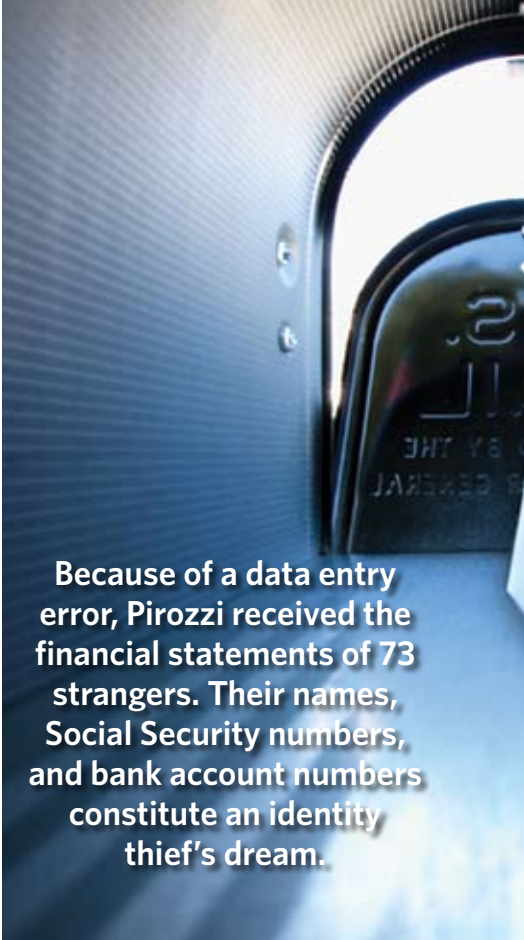
One day in May 2004, Ryan Pirozzi of Edina, Minn., opened his mailbox and found more than a dozen bank statements inside. All were made out to his address. All contained sensitive financial information about various accounts. None of the accounts, however, was his.

Because of a data entry error made by a clerk at the processing center of Wachovia Corp., a large bank headquartered in the southeastern United States, over the course of at least nine months, Pirozzi received the financial statements of 73 strangers, all of whom had had escrow accounts with this bank. All of these people, like Pirozzi, bought real estate through the Walker Title and Escrow Co. headquartered in Fairfax, Va. Their names, Social Security numbers, and bank account numbers constitute an identity thief's dream.

Then, during January of 2005, Piro-

zzi began receiving completed 1099 tax forms belonging to many of these people. After inquiries from a reporter for *The Washington Post*, both Wachovia and the Walker Co. began investigating the problem. The reporter found that many people who purchased a condominium unit at the Broadway in Falls Church, Va., were affected. These home buyers were given a discount for using the developers' preferred choice, Walker, to close on the purchase of their condominium units. In order to secure a condominium unit in the new building, prospective home buyers made deposits that were held in an escrow account at Wachovia.

According to Beth Givens, director of the Privacy Rights Clearinghouse headquartered in San Diego and quoted in the *Washington Post* article, "this case demonstrates that identity theft doesn't



Because of a data entry error, Pirozzi received the financial statements of 73 strangers. Their names, Social Security numbers, and bank account numbers constitute an identity thief's dream.

berry Pop-Tart™ toaster pastries. Thanks to our associates in the distribution centers and our drivers on the road, merchandise arrived quickly."

Federal Express

FedEx InSight is a real-time computer system that permits Federal Express' business customers to go online to obtain up-to-date information on all of their Federal Express cargo information. This includes outgoing, incoming, and third-party shipments. For example, John Smith might buy a gift from Amazon.com for Mary Jones and want to find out if the gift has already been delivered to Mary's home.

The business customer can tailor views and drill down into freight information, including shipping date, weight, contents, expected delivery date, and related shipments. Customers can even request e-mail notifications of in-transit events, such as attempted deliveries and delays at customs and elsewhere.

InSight links shipper and receiver data on shipping bills with entries in a database of registered InSight customers. The linking software, developed by Trillium Software, is able to recognize, interpret, and match customer names and address information. The challenge in matching records wasn't with the records of outgoing shippers, who could be easily identified by their account number. The real challenge was to link the intended shipment recipients to customers in the InSight database.

The process, of course, required accurate names and addresses. The addresses on the Federal Express bills tend not to be standardized and to be fraught with errors, omissions, and other

anomalies. The bills also contain a lot of extraneous information such as parts numbers, stock-keeping units, signature requirements, shipping contents, and delivery instructions. These items make it harder to extract the required name and address from the bill. The system Trillium Software developed successfully met all of these challenges and was able to identify and resolve matches in less than one second, processing as many as 500,000 records per hour.

Albertsons Inc. (and RxHub)

Albertson's is concerned with the safety of the customers buying prescription drugs at its 1,900 pharmacies. It's crucial that Albertson's correctly identify all such customers. Albertson's needs an up-to-date patient medication history on each of its customers to prevent a new prescription from causing an adverse reaction to a drug a customer might already be taking. Here, Albertson's is concerned about real-time recognition—understanding at the point of service exactly who is the customer at the pharmacy counter.

For example, Mary Smith may have a prescription at the Albertson's store near her office, but she needs to refill it at another Albertson's, the one closer to her home. The pharmacist at the Albertson's near Mary's residence needs to know immediately what other medication Mary is taking. After all, there's at least one high-profile lawsuit per year against a pharmacy that results in at least a million-dollar award. Given this concern, the return on investment (ROI) for the solution comes pretty rapidly.

In addition to health and safety, there are also issues involving the coverage of the prescription drug portion of the customer's



DREAMSTIME

always stem from people being careless with their financial information; the institutions that people trust with that information can be just as negligent. Although the worst didn't happen here, information gleaned from misdirected mail can wind up on the black market, sold to the highest bidder."

There have been instances, Givens said, in which mail processing systems misfire and match each address with a name that's one off from the correct name. In those situations, she said, hundreds or even thousands of pieces of mail can go to the wrong address. But those kinds of mistakes are usually noticed and corrected quickly.

The *Washington Post* article also quoted Chris Hoofnagle, associate director of the Electronic Privacy Information Center, as saying: "It should be rather obvious

when a bank sends 20 statements to the same address that there's a problem. But small errors can be magnified when you're dealing with very large institutions. This is not your neighborhood bank."

The article also reported that Mr. Hoofnagle "said it would not be difficult for Wachovia to put safeguards in place to catch this kind of error before large numbers of statements get mailed to the wrong person." The article did not provide the specifics about such safeguards.

Finally, one day in January 2005, a strange thing happened. Mr. Pirozzi went to his mailbox and discovered an envelope from Wachovia with his address and his name. It contained his completed 1099 tax form for 2004. "That," Pirozzi said, "was the first piece of correspondence we received from [Wachovia] that was actually for us."

DecisionMaker™ Rocks

**22 states and 19 vehicle types for a new line of business;
Up and running in 7 months!
Ask us how we did it !**



"Speed to Market" and product flexibility are two of the key characteristics of DecisionMaker™ Rating.

The DecisionMaker Suite has all of the modules necessary to efficiently manage your P&C lines of business from quote to claims.

The DecisionMaker modules can be licensed individually and integrated with disparate applications via SOA or deployed as an entire suite for straight through processing.



Software Solutions for
Smart Decision Making™

Regional Sales: Honolulu (800) 836.6057 / West (877) 658.7458 / East (800) 947.4037

health insurance. Albertson's has responded to both the safety and cost problems by deploying Initiate Identity Hub™ software to first identify and resolve duplication, followed by implementing real-time access to patient profile data and prescription history throughout all its stores. This allows for a complete, real-time view of pharmacy-related information for its customers concerning (1) the medications that are covered, (2) the amount of the deductible, and (3) the amount of the co-payment at the point of service to enable better drug utilization reviews and enhance patient safety.

RxHub provides a similar benefit at the point of service for health care providers accessing information from multiple pharmacy organizations. RxHub maintains records on 150 million patients/customers in the United States from a consortium of four pharmacy benefits managers (PBM). Basically, everybody's benefits from pharmacy usage are encapsulated into a few different vendors. RxHub takes the sum total of those vendors and brings them together into a consortium. For example, if I have benefits through both my place of employment and my wife's place of employment, the physician can see all those in one place and use the best benefits available to me as a patient.

The prominence of data quality issues in corporate governance has become even greater with enactment of the Sarbanes-Oxley legislation.

Even if my prescriptions have been filled across different PBMs, my doctor's office is able to view my complete prescription history as I come in. This is real-time recognition in its ultimate form. Because there is a consortium of PBMs, RxHub can't access the data from the different members until it's asked for an individual patient's medical history. Then, in real time, RxHub identifies and links the records from the appropriate source files and consolidates the information on the customer for the doctor. RxHub is able to complete a search for an individual's medical history in under 1/4 second.

Choice Hotels International

Choice Hotels International had built a data warehouse consisting entirely of its loyalty program users in order to analyze its best customers. Choice assumed that its loyalty program users were its best customers. Moreover, in the past, Choice could uniquely identify only customers who were in its loyalty program.

Then, Choice hired Initiate Systems Inc. to analyze its data. Initiate Systems discovered that (1) only 10 percent of Choice's customers ever used a loyalty number, and (2) only 30 percent of Choice's best customers (those who had at least two stays during a three-month period) used a loyalty number. So Choice's data warehouse contained only a small portion of its best customers.

Once Initiate Systems was able to implement software that uniquely identified Choice customers who had never used a unique identifier, it was able to give Choice a clearer picture of its true best customers. By using Initiate's solution, Initiate Identity Hub™ software, Choice now stores data on all of its customers in its data warehouse, not just the 10 percent who are members of its loyalty program.

For example, a customer might have 17 stays during a calendar year at four different Choice Hotels and never use a loyalty number. Because the hotels are franchised, they all have their own information systems. So the information sent to Choice's data center originates in different source systems and in different formats. Initiate Systems' software is able to integrate these data within the data warehouse by uniquely identifying Choice's customers across these disparate source systems. ●

THOMAS N. HERZOG is chief actuary with the Federal Housing Administration in Washington. FRITZ J. SCHEUREN is a general manager with NORC in Washington. WILLIAM E. WINKLER is a principal researcher at the U.S. Census Bureau in

PROFESSIONAL.

COMPETENT.

A COMPANY
YOU CAN TRUST.



STEWART SEARCH
ACTUARIAL RECRUITMENT

PLACING ACTUARIES WORLDWIDE
LIFE, HEALTH, PROPERTY & CASUALTY,
PENSION, ANNUITIES

www.stewartsearch.com 888-JOB-OPEN