

The Penalties of Poor Data

Tyler Anderson IT Director

Have you ever considered how much time and resources your organization spends on correcting, fixing and analyzing corrupted or erroneous data? What about the cost of delayed information exchange or lost revenue due to misplaced data or incorrect input? Evaluating data and determining errors is a time consuming process, not to mention the time needed to correct them. In a time of decreased budgets, some organizations may not have the resources for such projects and may not even be aware of the problem. Others may be spending all their time fixing problems leaving no time to work on preventing them. According to several leading data quality managers, the cost of poor data quality may be expressed as simple formula that equates into:

meta-data is definitional data that provides information about or documentation of other data managed within an application or environment, basically data that controls data.

Cost of Poor Data Quality (COPDQ) = Cost to Prevent Errors + Cost to Correct Errors + Cost to “Make Good” for the Customer.

Let’s use an example to illustrate this point. Say you have data entry staff working on cleaning up a data base list of addresses

1. A staff person spends about 20% of their time each day on this task
2. There are two people performing this operation (2 x 1.6 hours = 3.2 staff hours per day.)
3. Accounting tells you that these people earn \$12 per hour (payroll + benefits.)
4. Total annual hours of cleanup is 800 hours annually (3.2 staff hours x 250 annual working days.)
5. This means the annualized COPDQ to fix the addresses is \$9,600 (800 hours x \$12 per hour + benefits.)

This cost of the poor data quality extends far beyond the cost to fix it. It spreads through and across the business enterprise like a virus affecting systems from shipping and receiving to accounting and customer service. Eventually, your customers may lose patience with you, and you may lose their business.

The Data Warehousing Institute estimates that data quality problems currently cost U.S. businesses over \$600 billion annually. They traced this figure back to things such as unnecessary printing costs, staff overhead, and mailing costs. Their reporting goes on to estimate that at least 2% of all data is corrupted each month due to name changes, divorces,



deaths and people moving. These figures came from 647 respondents and was cross analyzed against Dun & Bradstreet data.

Even with these figures to guide us, it is still very difficult to use metrics to determine the cost of poor data quality and its effects on your organization. This is because making the mistake may be too distant from recognizing the mistake. Errors are very hard to repair, especially when systems extend far across the enterprise, and the final impact is very unpredictable.

So with that said, how can you figure out what is necessary for your organization? Chuck Kelley of Excellence In Data, Inc. says that *most* companies can get away with 20-30% “erred” data. He goes on to state that companies must carefully weigh the costs of their return on investment before installing a total data quality system.

The first step in cleaning your data is to figure out what might prevent you from achieving such a goal. DM-Direct suggests there are five obstacles to achieving true data quality.

Data Is Not Easily Accessible

Many organizations have multiple databases extending far across the enterprise. For example, accounts receivable may have a database of all outstanding invoices, while accounts payable maintains a completely separate database for vendors. The marketing department may have a database for prospects, while sales is maintaining an ACT or Goldmine database for its current “hot” prospects or leads. Inside sales or Customer Service may be generating orders or taking calls using a completely different system from all of the above. These databases may be strewn across several servers in several facilities—or even worse, on a person’s local workstation—making them inaccessible to the rest of the organization. One solution is to connect these databases to allow information to be shared and updated simultaneously using a data pipeline. Another is to have one enterprise wide system that contains all this information in a data super warehouse.

Data Is Not Well Documented

Organizations collect, store and maintain endless terabytes of data, but don’t usually document where it came from or what it should be used for. A guidebook of metadata must be developed for the users to know how to interpret the data and how to use it. For example if you have a field called comments under the company name, are these comments for the entire company or for an individual contact or for a certain job related issue? The solution here is to maintain field descriptors or data about what fields are used for what.

Data Is Poorly Managed

Collecting and maintaining data are two different animals. Collecting data can be analogized to the raw building materials for a house while maintaining it is like upkeep on that house. In order for data to be a corporate asset it must be managed as such with these



two initiatives kept separate at all times. For example, a university may have multiple ways of contacting past, current, and prospective students. The students and alumni may want to receive communication via e-mail, fax, phone or mail. They may also want to receive information for one type in a certain format, while receiving other types via a different format. The incoming call center must make sure this information gets to the admissions marketing department. When data is managed strategically, it is made actionable.

Data Is Not Integrated

Very few companies integrate data systems across the platform well. Have you ever hooked up cable with a local service provider and months, even a year later received a card mailer requesting you sign up for cable service with the same company? Problems like this drive up marketing costs, skew marketing results, cost the end user (client) more and irritate customers. Many CRM (Customer Relationship Management) systems have failed because companies don't recognize the customer data that feeds these applications. The solution here is simple: integrate your data into one data warehouse storage facility across the entire enterprise.

Data Is Not Valued

This is probably one of the most detrimental contributors to data quality. Data advocacy is a company wide cultural value that brings together two divergent objectives: data management and data use.

The inherent problem is that two completely different groups are managing and using this data. Business executives use the data, while technology professionals manage it. Often more than not, these two groups don't communicate effectively enough thereby limiting the value of the data within their systems. Business executives just want to get resultant information out of the systems for analysis to reach company goals. On the other hand, IT professionals may not know the companies overall business goals and how to turn all the raw data into something manageable for the business executives.

To bridge this divide, companies must select a "champion" to oversee the management of the companies directives on data quality. This person must bridge the gap between management and IT professionals to help them understand the needs, limitations, and expectations of the data being managed. This person must also be able to teach and educate personnel from the top to the bottom on how valuable data is and how costly erred data will be for the company. The use of financial numbers related to costs and savings should be used whenever possible to prove the value of data quality.

Tyler Anderson, IT Director, immedia

