



# “Poor-Quality Data... Can Your Company Afford the Risk?”

## Noteworthy



Claudia Imhoff, Ph.D., President and Founder of Intelligent Solutions, Inc., is a thought leader, visionary, and practitioner in the rapidly growing

fields of business intelligence and customer focused-strategy. Dr. Imhoff founded Intelligent Solutions, Inc. ([www.IntelSols.com](http://www.IntelSols.com)), a well respected Business Intelligence and CRM consulting and education firm in 1992. Her company has successfully implemented over 150 Corporate Information Factory architectures in all industry areas.

A popular and dynamic speaker and internationally recognized expert on analytical CRM, business intelligence, and the infrastructure to support these initiatives, Dr. Imhoff delivers keynote addresses at conferences sponsored by software companies and their user groups, The Data Warehousing Institute, The Economist, COMDEX, and many international organizations. She has appeared repeatedly on World Business Review, Microsoft's Getting Results programs, and web casts sponsored by the B-EYE-Network, Better Management, and several technology vendors. Dr. Imhoff has co-authored five highly-regarded and popular books on these subjects, has written numerous articles (totaling more than 80) for technical and business magazines, and has a popular blog at [www.B-EYE-Network.com](http://www.B-EYE-Network.com). She has served on the Board of Advisors for DAMA International and was chosen by the DAMA organizations to receive the 1999 and 2005 Individual Achievement Awards. She is an advisor and a faculty member for The Data Warehousing Institute and serves as an advisor for several technology and commercial companies. She is also a member of the Advisory Board of the Daniels School of Business at the University of Denver.

By Claudia Imhoff, PhD, President, Intelligent Solutions, Inc.

### 1. Let's start by defining what you mean by poor-quality data.

It's an insidious problem, but also one that is difficult to define. It's more easily calibrated in terms of what companies believe they have lost – for example, real revenues or real savings. That's the prevailing definition of what poor data does. Poor-quality data, of course, is data that is inaccurate, incomplete, and misleading. And that leads to bad business decision-making.

### 2. No company, of course, wants poor-quality data. So... how do companies find themselves in situations where they do have poor-quality data?

The problem that companies have is that data is captured in multiple operational systems, each set up to do a specific task, such as enter orders. What if a customer is entered twice? Then you have a fractured customer record. This could be because the person doing the entering has the incentive of entering orders as quickly as possible. And if it's quicker to enter a customer anew rather than look up the customer record, then the company ends up with multiple entries for the same customer... some of which may vary slightly. That's just one example of poor quality data. Now multiply this by the number of operational systems that contain customer data – billing, sales force automation, general ledger, etc. – and you can see the magnitude of the problem just for customer data.

Many companies have situations like this, where incentives are way out of whack in terms of being able to foster a good data quality environment.

Additionally, systems might be used in ways that the person entering the data doesn't really understand. For example, they might have to enter standard industry codes. That has nothing to do with the order entry clerk's ability to fill the order. But what if that information is critical downstream to other associates trying to analyze what types of companies are buying from us? Who should we target for our marketing campaigns?

The source of poor data quality isn't malicious or intentional; it's just that people want to enter the data expeditiously, and quality isn't always rewarded.

### 3. What is the fallout of operating with poor-quality data? The risks?

As I mentioned earlier, bad decision making is the biggest risk. To continue the order entry example, the marketing person down the line gets bad intelligence, incomplete views of customers, and might even think a customer is unprofitable when really it is quite profitable. They simply can't get the return they want on their marketing campaign because without the right data, you can only have mistaken analysis and erroneous decisions.

You've never seen your business like this before.

**Teradata**  
a division of  NCR

#### **4. Companies can't stop functioning – take a break, so to speak – to clean up their data. So, what are the processes companies can follow to improve their data while still managing day-to-day operations?**

It's like changing the tires on a moving car! Companies can't suddenly – even temporarily – stop the influx of data, but they can put in processes to stop influx of bad data.

First, I'd look at the incentive packages of employees who handle critical data. Are they being incented properly and appropriately? It can be challenging, but paying based not just on amount of data entered but also quality of that data entry, is a big place for companies to start improving the quality of the data right as it enters the company.

Next, it's essential to improve existing data through technology and data quality tools. These tools have come a long way and are very sophisticated in terms of accurately matching fractured customer data up. Existing data can be almost completely cleansed, providing the company with the critical "single view of the customer."

Another step to take is to determine: who is responsible for bad – or good quality – data? The truth is that responsibility for data quality crosses all sorts of enterprise boundaries. Quality data doesn't belong to any one single department and improving poor quality data involves multiple departments, organizations, and individuals in the corporation.

So, a movement is afoot that I think is a good one – some companies are starting to create organizations, or cross-enterprise teams, of data stewards who do have not only the responsibility but also the authority to improve business processes so that the same data quality problems don't keep cropping up over and over.

#### **5. When is data "good enough?" In other words, what parameters can help a company know it's moved from poor-quality to high-quality data?**

If you can't measure it, you can't improve it. So a fundamental part of improving data quality is to be able to measure data quality. Until you have some type of baseline metric, you don't even know where you are. Now, is data ever good enough? If the person doing analysis of the data knows that the data is consistently ten percent off for whatever reason, then the analysis can be adjusted or accommodated for that. From an analysis standpoint, that's perfectly good data. You can adjust for imperfections if they are consistent and you know what they are. If you can consistently say data is such and such, then you can use it.

Teradata and NCR are registered trademarks of NCR Corporation. NCR continually enhances products as new technologies and components become available. NCR, therefore, reserves the right to change specifications without prior notice. All features, functions, and operations described herein may not be marketed in all parts of the world. Consult your Teradata representative or [Teradata.com](http://Teradata.com) for more information. No part of this publication may be reprinted or otherwise reproduced without permission from Teradata.

© 2006 NCR Corporation Dayton, OH U.S.A. Produced in USA All Rights Reserved.

But if you measure data quality and it is too low – let's say it's 80 percent accurate and you need to get to 90 percent accuracy – then that's where things can get complex. It helps to work through that complexity if IT and business users can collaborate to discover where the bad data is coming from. Where are the problems? Is it fixable? Is there some fundamental process that is missing? Perhaps retraining is needed? Maybe redirected incentives? Can the technology be imposed on the problem to help correct it?

In some cases, it may be discovered that it's too difficult or costly to fix a particular problem – so it's on to the next one. Not all problems will fall into the "yes, it's worth the effort to fix it" column. But others will certainly be fixable and the company can get closer to that 90 percent accuracy goal.

#### **6. What role does data warehousing and other technology play in enabling companies to have high-quality data?**

Data warehousing is the first place that we bring fractured data from operational systems together. It's the first place where we try to integrate the data into a single source. Therefore, it is the messenger of, oh no, we have bad data. But – please – don't shoot the messenger! The messenger is simply pointing out that there are problems somewhere in the operational systems or processes or gathering of data. Unfortunately, many people, when they see poor quality results of their data analysis, blame the data warehouse. The data warehouse is not responsible for bad data. It's also not fair to tell the data warehouse team to fix all the data quality problems they encounter, because most – especially the insidious ones – were created by the data being inappropriately entered in the first place. It is much more cost-efficient to fix the problems where they occur.

#### **7. What are the benefits of higher quality data in terms of Customer Relationship Management (CRM)?**

Data quality enables business decision making to be performed with confidence. If we have better quality data as it's gathered, and the quality is managed with data quality tools, then in the long term we can get closer to a consistent, 360-degree view of the customer, and make strategic CRM decisions with greater accuracy. We can be assured that we've created CRM programs that truly fit the customer, because we have an accurate reflection of the actual customer in our data. With that confidence of accuracy, we can do better analysis of why a customer is profitable or not, how to make the customer more profitable, how to grow the business.

It's important to remember, though, that data quality is never a one-time fix. It is a continual, enterprise-wide process that everyone must support.

You've never seen your business like this before.

**Teradata**  
a division of  NCR