

Foreword to

XML in Data Management: Understanding and Applying Them Together

by Peter Aiken and M. David Allen

An exercise popular in many training courses goes something like this. The class is asked to imagine a fine antique French desk, recently purchased for \$20,000. Atop the desk sits a brand new laptop computer, complete with all the bells and whistles, that cost \$3,000, and a diskette that cost a dollar. The diskette contains the only known list of the names and purchases of the organization's fifty largest customers. Now, the exercise goes, a fire has started and you can only save one of the three. Which do you save?

Naturally, almost everyone chooses to save the diskette. And they immediately recognize that it is not the diskette that is worth saving, but the data it contains. Almost everyone intuitively concurs that the data are worth more than \$20,000. The exercise illustrates a point that everyone knows but doesn't think too much about—that data are an extremely important asset. When they do think about it, everyone realizes that those with the best data craft the best strategies, make the best decisions, know more about customers, and keep costs down.

Data have always been important. Those with the best data win wars, define and execute the best strategies, make better decisions, and keep costs down. And, since data are central to the scientific method, the process so successfully used to learn more about our world, they have played a role in virtually every advance for thousands of years. There are any number of examples in which business organizations have exploited data. The most recent spin is named "data mining," though statisticians, market researchers and others have been uncovering hidden nuggets in data for generations. Virtually everyone would agree that data offer almost limitless potential.

At the same time, data are perhaps the organization's most underutilized asset. Consider some of the most-frequent complaints of knowledge workers, day-in and day-out:

- "I can't find the data I need to do my job."
- "I can't understand the data."
- "The data are wrong."
- "These data and those data don't agree."

Data Blueprint is a professional services company in Richmond that was spun off from Virginia Commonwealth University to capitalize on developed intellectual property. We create business value by applying innovative data (re)engineering/management techniques for recovering, understanding, and improving data assets. We also sponsor research that furthers the study of data and metadata, increasing the value of your data assets.





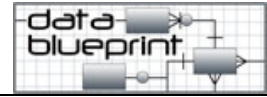
One need not work through the intricacies of IT architectures to understand this point. One need only read the newspaper. Poor quality data lie at the root of issues that capture our collective attention and will not let go. A partial list includes:

- Type A organs were mistakenly transplanted into seventeen year-old J sica Santill n, eventually killing her. And two other lives that could have been saved with these organs, were not.
- Corporate scandals that revolved around incorrect financial reporting dominated the news in late 2001 and 2002. Even when there is no suggestion of wrong-doing, hundreds of financial statements must be restated because errors are found.
- Some people weren't counted in the most recent census. Others were counted twice.
- Many speculate that, had the FBI and CIA shared data, they could have pieced together the clues and prevented the September 11 attacks.
- The contested Presidential election in 2000 was essentially a data issue—the data being the votes. One is tempted to dismiss the Florida situation as an aberration that can't happen again. But miscounted elections are all too common.
- In 1999 the US Military bombed the Chinese Embassy in Kosovo as the result of an out-of-date map.

Fortunately, most data quality issues don't make the front page. But virtually every organization is bedeviled by bad data—increasing cost, angering customers, compromising decisions, and making it more difficult to align departments. One study puts the costs to American business at over \$600B/year for *customer data alone*.

There are no magical answers, no silver bullets for addressing these questions and issues. Nor are they intractable. A vanguard of organizations have demonstrated that order-of-magnitude improvements in quality are possible. They have found that the technical challenges are relatively small compared to the organizational, social, and political challenges. But, as people and organizations “see the light,” start to think and manage differently, and apply themselves to the task, dramatic improvements follow.

XML is an important tool. It can help people find the data they need and understand what these data mean. In time, it can help organizations figure out which data are most important and how they use them. Further, data add most value not when they are stowed away in a database, but as they fly around the organization. And that is the “space” (I like to call it the “data exchange layer,” others may call it “messaging,” “communications,” ETL, etc.) where XML is brought to bear.



All this said, XML is only a tool. XML is an important complement to, not replacement for, well-defined, disciplined data management processes, and a well-conceived and executed data exchange layer.

Peter and David bring these points into focus. They define the yin and yang of XML and data management. They help separate the hype from the reality. And they equip a manager to make better choices moving forward.

Thomas C. Redman, Ph.D.

President, Navesink Consulting Group