tactical DATA QUALITY

How to improve data quality with a tight budget

In this e-book, you'll learn how to manage data quality efforts during an economic downturn and find out what trends are emerging in the data quality market. You'll also learn about common mistakes, how to avoid the pitfalls of poor data and how data quality tools and strategies can improve data quality.

TechTarget
The Technology Media ROI Experts

Managing data quality programs during a recession

Forgoing data quality management during a recession can cost you, according to experts. First, assess ROI for managing data quality programs and the impact and cost of poor data quality. **BY DAVID LOSHIN**

DURING UNCERTAIN ECONOMIC times, there is a certain amount of belt-tightening expected across the board, and the IT department is not immune to this. Yet before you grab the knife to start slashing the budget, it is worth considering that reducing the investment in any program or infrastructure that supports the organization's business needs is a measure that will not only diminish needed agility during poor economic times but will also slow the organization's competitiveness when times start to get better.

Often, data quality management is seen as a good practice, but most organizations do not have the discipline to integrate its value proposition holistically across the organizational value drivers, whether they are focused on revenue growth or operational cost containment. Therefore, a recession actually provides an excellent opportunity to assess two aspects of the relationship between data quality and the business. Companies can directly connect high-quality data to the organization's value drivers, weighted by the perception of existing economic trends.

DETERMINING DATA QUALITY'S IMPACT ON BUSINESS PROCESSES

The first aspect is identifying specific business processes that will be positively affected by high-quality data. Data quality may affect different business processes in different ways. A data quality analysis should incorporate a business impact assessment to identify and prioritize risks. Those business impacts associated with bad data can be categorized within four general categories for assessing

either the negative impacts suffered or the potential new opportunities for increased value resulting from improved data quality:

- **Revenue growth**, incorporating financial impacts such as decreased sales, higher cost to acquire new customers, and customer retention.

- **Cost reduction**, such as increased operating costs, reduction or delays in cash flow, and additional unnecessary charges.

- **Risk management** and confidence management, such as credit assessment, investment risks, competitive risk, capital investment and/or development, fraud and leakage, compliance risks, decreased organizational trust, low confidence in forecasting, inconsistent operational and management reporting, delayed or improper decisions, decreased customer, employee, or supplier satisfaction, or lowered market satisfaction.

- **Productivity impacts** such as increased workloads, decreased throughput, increased processing time, or decreased end-product quality.

Assessing the business impacts associated with data means working with the business consumers to understand their information needs and the corresponding data quality expectations. One can elicit information about the business impacts associated with data quality by asking probing questions such as these:

① What importance does data have in achieving the organization's business objectives?

② What data is critical to your business processes?

③ How confident are you in the accuracy of your data?

④ What changes to the data can improve business process performance?

⑤ In which aspects of data improvement should the company be investing, and in what time frame?

Any significant data issues that will affect the business are likely to be revealed during this process, and this provides you with the basis for further researching documented business issues and connecting them to any type of data flaw. It will provide a connection between data quality improvement and a potential for increase in value. At the same time, this provides an opportunity to reinforce conformance with business data quality expectations by validating data quality rules and the corresponding thresholds for acceptability.

This leads into the second aspect of data quality management: monitoring the level of efficiency of the data governance and data stewardship activities. As a data quality program matures, the management of issues transitions from a reactive environment to a proactive one, and this can be scored in relation to continuous monitoring of the quality of data. In the optimal environment, the data stewards allocate time to address the most critical issues as they are identified early in the processing streams. Less efficient organizations have stewards reacting to issues at their manifestation point, at which time these issues may have already caused significant business repercussions.

THE IMPORTANCE OF DESIGNING A DATA QUALITY SCORECARD

Therefore, organizations that inspect, monitor and measure the performance of data quality initiatives on an ongoing basis, across all processing streams, can then populate a data quality scorecard reflecting the effectiveness of the program and the efficiency of its staff. Together, these two aspects reflect the value of the program and the way that it has been implemented. Focusing on both of these aspects provides a number of valuable benefits:

- It can demonstrate the value proposition for maintaining the effort, even in the face of economic stress.

- It can provide long-term justification for continued funding and growth of data quality management as the recession ends.

- It can help identify additional areas with an acute data quality improvement need that can help support the organization's survival during a recession.

- It will demonstrate an example of proactive value management to other organizations.

On the other hand, it may turn out that this data quality assessment will show that the organization does not get a reasonable return on its data quality management investment. In this case, it provides an opportunity to reduce operating costs associated with the areas of missed expectations. While this is unlikely, it does demonstrate a level of accountability that should pervade all management activities. It is more likely, however, that this process will only strengthen the view that a data quality management program is fundamental to the ultimate success of the business. ■

David Loshin is the president of Knowledge Integrity, Inc. He can be reached via knowledge-integrity.com.

Trends in the data quality market

The data quality software market continues to grow as tools for non-IT users gain in popularity. Find out what trends are shaping the data quality industry and what vendors and tools have the biggest share of the data quality market. **BY JEFF KELLY**

NOT EVEN A WORLDWIDE recession could hold down the data quality software and tools market, which continues to grow as more organizations recognize the importance of quality data to master data management, business intelligence and other information management initiatives, according to Gartner's latest Magic Quadrant report.

More organizations are also coming to the realization that poor data quality is a pervasive issue that puts a drag on performance, revenue and profits, said Andreas Bitterer, an analyst with the Stamford, Conn.-based analyst firm and co-author of the report, along with Gartner analyst Ted Friedman.

"Companies are realizing bad data is having a negative impact on performance," Bitterer said. "There's not one company on this planet that doesn't have a data quality problem."

The data quality and data integration markets also continued to converge in 2008, a trend Gartner identified in last year's data quality Magic Quadrant report, though not as quickly as Bitterer expected. It will still be several years before data quality and data integration software are routinely bundled in a single platform, he said, but the market is headed in that direction.

"Whenever you move data [be it through ETL, data federation or other methods] while you have [the data] in your hands, you should make sure whatever arrives at the other end is consistent and complete," Bitterer said.

The data quality market, which Gartner estimates stood at between $400 million and $500 million as of the end of last year, is largely divided

between big, incumbent vendors that offer a wide breadth of functionality —including data profiling, standardization, cleansing, matching and enrichment—and small niche vendors with more targeted but less expansive capabilities, according to the report.

The large, incumbent vendors, notably SAP BusinessObjects and IBM, "increasingly focus on data quality capabilities as complementary to various components of their portfolios," Bitterer and Friedman wrote in the report. "While they sell data quality tools in a standalone manner [as individual products], these tools are increasingly sold as part of a larger transaction involving related products [such as data integration tools and MDM solutions]."

Data quality analysis tools for non-IT workers, like data quality dashboards and visualization applications, also continued to develop as organizations and vendors alike increasingly recognize data quality as a business problem, not just an IT problem, Bitterer said.

DATAFLUX SOFTWARE AND TOOLS TOP DATA QUALITY RANKINGS

DataFlux, a subsidiary of SAS Institute, based in Cary, N.C., topped Gartner's data quality vendor rankings this year. The vendor's broad data quality capabilities are easily integrated with other applications and relatively easy to use, especially for non-IT workers, the report found.

"With its 1,200 customers, DataFlux has become the enterprise-wide data quality standard in many large accounts," Bitterer and Friedman wrote. "The company has one of the highest ratios of reinvesting revenue in R&D and enjoys a maintenance renewal rate of over 95%."

Joining DataFlux in the leaders' quadrant were Trillium Software, IBM, Informatica and SAP BusinessObjects. Gartner's Magic Quadrant methodology places vendors that meet its inclusion criteria into one of four quadrants based on "completeness of vision" and "ability to execute." The quadrants are niche vendors, visionaries, challengers and leaders.

Trillium Software was cited for its "diversity of use cases, including those within BI activities, MDM solutions and in support of data governance programs," while Informatica landed in the leaders' quadrant on the strength of its data profiling functionality and domain-agnostic data parsing, standardization and matching capabilities, according to the report.

The niche players' quadrant was also crowded this year, with six vendors sharing the space. They are DataLever, Uniserv, Innovative Systems, DataMentors, Netrics and Datactics. While these vendors, among

other smaller data quality vendors, offer mature if narrow data quality capabilities, they are likely to struggle to gain market share from their large, incumbent competitors, according to the report.

"With the increasing trend toward embedding data quality capabilities in business applications, data integration tools and other software offerings from larger vendors, these small competitors will face significant challenges as they attempt to survive and grow," the report stated.

Other vendors meeting Gartner's evaluation requirements included visionaries Human Inference and Datanomic, and Pitney Bowes Business Insight, the lone challenger.

DATA QUALITY NOT JUST AN IT PROBLEM ANYMORE

Data quality is as much an IT issue as it is a business issue, Bitterer said, so companies evaluating data quality vendors should heavily weigh the software's ease-of-use, especially for non-IT workers.

"The content is owned by the business," Bitterer said, and in most cases the business is the group that understands the nuances of the data best, those aspects most critical to data quality.

Therefore, the business must play a significant role in any data quality initiative, with tools to match their abilities and understanding. Vendors have largely made this connection already, with most offering data quality dashboards and other visualization tools to help non-IT users monitor and manage data quality projects, Bitterer said. And many organizations have already adopted them.

Bitterer also recommends that companies consider data quality vendors' location, as many vendors are stronger in European markets than in the U.S. and vice versa. Data quality is often affected by regional variations—the way an address is formatted, for example—and a vendor located in its customer's geographic area is likely to understand those differences better, he said.

Finally, Bitterer said that all companies evaluating data quality vendors should consider both the large, incumbent vendors as well as the smaller, niche players. In some cases, companies with more targeted data quality issues can get the functionality they need from the smaller vendors at a much better price than from their larger counterparts.

"The companies in the upper right corner [of the Magic Quadrant] are a lot more expensive than the ones in the lower left corner," Bitterer said. "Not everybody needs the 38-ton truck." ■

Jeff Kelly is a news editor at SearchDataManagement.com

Avoiding data quality pitfalls

Less than a third have deployed data quality tools enterprise-wide, according to Gartner, Inc. Find out why some groups haven't embarked on a data quality project and learn how poor data is hurting businesses. **BY JEFF KELLY**

HIGHLY ACCURATE, up-to-date data—especially customer data—is one of the keys to maintaining strong customer relationships and ultimately growing revenue.

But, while most organizations acknowledge the importance of data quality, only around half of all companies have actually deployed data quality tools or started data quality initiatives, according to Ted Friedman, an analyst with Stamford, Conn.-based Gartner Inc.

Even among organizations that do use data quality tools—either commercial or homegrown—less than a third have deployed the tools enterprise-wide, according to Gartner.

The consequences are not trivial. Poor or siloed data can result in missed cross-sell and up-sell opportunities and can even alienate customers who have come to expect personalized interactions.

"Especially in the under-40 demographic, customers do expect a high level of customization/personalization from companies—and this puts pressure on companies to deliver or risk losing their existing customers," Leslie Ament, managing partner at Lexington, Mass.-based Hypatia Research, said in an email interview.

So why do so few organizations use data quality tools for customer data enterprise-wide? The reason, according to some, is that most companies collect and store customer data in numerous data sources spread throughout the organization with no way to connect them.

Put another way, lacking a single view of the customer through a master data management (MDM) system or customer data integration (CDI) initiative, organizations lack any realistic way of applying data quality tools enterprise-wide. Their only alternative is to tackle customer data quality one department or data-

base at a time.

"Many larger retailers have upwards of 10 different databases with different schema for collecting customer data," Ament said. "Standardizing and normalizing this information is akin to having root canal surgery at the dentist."

Navin Sharma, director of product management for global data quality at Pitney Bowes Group 1 Software, which recently released an enhanced version of its customer data quality platform, agreed. "Even when we talk to our customers, many of them are deploying our capabilities, but for specific needs [such as for marketing or compliance issues]," Sharma said.

"The issue [is] there is disparity in how data is stored and spread across the enterprise," he said. "The fact of the matter is most organizations still struggle to maintain that single view of the customer. From that perspective, the maturity as well as the adoption [of enterprise-wide data quality] is still very low."

Gartner's Friedman, in a 2008 Gartner report, said companies need to start thinking about data quality more broadly and in the context of enterprise information management and MDM initiatives. In fact, data quality is an integral part of any MDM program. With a single view of customer data achieved, the thinking goes, applying data quality standards to it is a much simpler task than attacking the problem one database or data warehouse at a time.

But lacking a full-blown MDM initiative, Friedman wrote, companies should still conduct an inventory of all the data quality tools being used through the enterprise and identify "ones that can be used broadly across the business to reinforce a uniform approach to data quality, as well as to reduce procurement and maintenance costs."

In some instances, companies with simple customer data quality needs can even get started with tools embedded in their existing applications, Friedman said in an interview. Most business intelligence applications include data transformation capabilities that can reconcile customer names, for example. But most organizations will need to invest in more specialized data quality tools for sophisticated tasks like data parsing and standardization.

Sharma, for one, predicts that adoption of customer data quality tools and their implementation across the enterprise, while still lagging, will eventually gain momentum as more and more companies recognize the advantage it can provide over competitors.

As Hypatia's Ament points out: "Using data to understand and respond to customers can make a huge difference in a crowded marketplace." ■

Identifying data quality problems with a data quality assessment

Data quality problems can't be fixed until they are identified. In this Q/A, Arkady Maydanchik reveals how a data quality assessment can help pinpoint data quality problems. **BY JEFF KELLY**

EDITOR'S NOTE: The following Q/A is an edited transcript based on a previously recorded data quality podcast on SearchDataManagement.com. Please contact editor@searchdatamanagement.com with any questions.

WHAT'S THE POINT of analyzing incorrect, out-of-date or mislabeled data? Actually, that's a trick question. There really is no point, which is why achieving comprehensive data quality—that is, ensuring the accuracy, reliability and effectiveness of data—and overcoming data quality problems is so critical to business success.

But data quality is still an emerging field, one often overlooked by companies and organizations. To help understand how to start a data quality program, SearchDataManagement.com caught up with **Arkady Maydanchik**, a member of the Data Quality Group and author of the recent book, *Data Quality Assessment.*

But before we get started, here's a little more background on our guest speaker. He is a recognized practitioner, author and educator in the field of data quality and information integration. His data quality management methodology was used to provide data quality services to numerous Fortune 500 companies. Arkady is a frequent speaker at various conferences and seminars, author of the aforementioned *Data Quality Assessment*, and a contributor to many journals and online publications.

So Arkady, you said that data quality assessment is the starting point for any data quality program. Can you explain to our listeners what a data quality assessment is and why it is

so important when starting out on any data quality program?

Sure. Basically, we've been talking about data quality for probably the last 10 or 15 years, and back 10 years ago, it wasn't a very popular topic. Now it seems like it is a very popular topic. Everybody wants to talk about it; everybody says, "Our data is really bad." A lot of people are saying, "Hey, let's try to do something about it. Let's try to fix the problems. Let's try to improve the processes."

The reality is you can't fix the problem until you know what it is. That's actually one of the reasons why, even though we've been talking about data quality for 10 or 12 years, not much has really been done. Frankly, data quality deteriorated over the last 10 years quite a bit.

Data quality assessment is the process of taking your existing data and systematically going through it and identifying what's wrong with it, [figuring out] where the data problems are, and then connecting that to your real business process and understanding how the data problems impact your processes and what is the cost of your data quality problems. Once you have done that, you can do a lot of things. Then you can say, "OK, we understand now how much it costs, so now we can possibly do a return on investment (ROI) analysis and figure out if it makes sense for us to invest money into fixing some problems." Once you know where the problems are, you can try to analyze where they came from—do a root-cause analysis and try to prevent problems from happening in the future. All this relies on the fact that you know where your problems are and what they cost your organization.

> *"Frankly, data quality deteriorated over the last 10 years quite a bit."*
>
> *—ARKADY MAYDANCHIK*

It's an important step for both the technical reasons and the political reasons. Political reasons are also important. Most organizations today don't really have data quality departments. Most of them are taking the first step, and I oftentimes have people come to me and say, "Hey, you know we think our data is bad; we want to start a data quality management program, but we need some way of selling it to our management." They say, "Hey, is there any template, is there any way we can sell this to

our management and tell them, prove to them, that we need a data quality management program?" The truth is you can't just go and say, "Hey, everybody is talking about it, and there are many reasons why you need to do data quality management." What you need to do is show how it affects your specific organization. Until you have that knowledge, you really can't get any budget or start any programs. So, both on the side of starting a program and on the side of getting anything done, you need to have a specific, systematic knowledge of what is wrong with your data and how it impacts your business.

You've also talked about data quality rules. What role do they play, and why are they so crucial to the outcome of any data quality program, as you wrote in your recent book?

Now, we said we want to find the problems with our data. Now, fundamentally, there are three approaches you can follow. One is what we'll call a complete manual validation. Basically, you take every piece of data and compare it with some trusted source. Maybe it's a paper file or maybe it's checking with some real people who have the knowledge of that information. Now, that is totally impractical because our databases are made of millions or billions of data pieces, so we really cannot do it.

Another way is to take the page from the common quality assessment. Data quality assessment is something reasonably new. But quality assessment in many industries is always done. Automobiles go through quality assessment, vacuum cleaners, airplanes, every product that is made out there goes through quality assessment. The other approach is to say, "Let's try and pick up a sample of the data. So we have millions of records. Instead of doing all of them, we'll take a few hundred and analyze them. Based on their quality, we'll try and project how good the quality of the rest of the data is." Well, that also doesn't work

"You need to have a specific, systematic knowledge of what is wrong with your data and how it impacts your business."

—ARKADY MAYDANCHIK

because it can give us an idea of how bad the data is, but it can never point us to any specific data elements that are wrong because we'll only know that about our sample.

The beauty of the data is it describes real, concrete objects. It describes people. It describes products. It describes customers. It describes processes. And they have many attributes and characteristics that are all intertwined and interrelated. And so the attributes of the data itself are also tied by millions of different relationships. Those are the constraints on the valid value. Those are the relationships across data elements. The relationships in the order certain events can occur, on the condition in which events can occur, the timing of when certain measurements are made. So there are hundreds and hundreds and hundreds and for bigger databases thousands, of different constraints and those are the ones we call data quality rules. Constraints that we can apply to the data—and anytime data violates those constraints then there's something wrong with it. Now, the beauty of this is that those constraints can be implemented in computer programs. We can run all of our data no matter how large our database. We can run all of the data through these constraints, and basically, within the reasonably short time frame of the project, we can find all of the inconsistencies. It's not going to be 100%, but we can find probably 95%, 98% of all potential data problems.

So, data quality rules are at the heart of this because they are a tool, they're the main tool of a data quality professional.

How do you ensure that your data quality assessment is comprehensive and complete?

That's a very good question because it's quite easy to design a bunch of rules, but the challenge, the difficulty is making sure we really did find all of the problems, and also the challenge is to make sure the discrepancies we find are true problems. The kind of analogy I usually give is if I was asked to be an umpire in a Major League Baseball (MLB) game. Now, if you look at the rule book, it has 136 or 137 rules. Now let's say I learn all but 10% of them, so I learn 120 rules. The odds that I'm going to cause a riot during the games because I missed some play are pretty high. Just knowing 90% of the rules doesn't really do the job.

The same problem with the assessment; because there are so many different constraints on the data you can think of, it's not hard to quickly write a short list of scores of hundreds of them. But how do we know we're finding them all? Here is where

the other approach, assembling, comes to help. What we do is first come up with all these rules, and there's a very systematic approach to that in my book. I dedicate half of the book to systematic processes of how to identify all of the different rules. Then what we do is take a sample of the data and we use the true data expert, somebody who has a knowledge of and access to the data sources. And we ask that expert to validate a sample of the data, and then we compare the findings with the results of the rules. Basically, the objective is, even though we're using the rules, we want to find all the same problems that the data expert would've found using the objective trusted data approach.

So, once we match what the data expert finds with what the rules find, we'll see whether the rules have shortcomings. There might be some errors that the expert found that our rules didn't. Well, then we should try to find what other type of constraints and rules we can design that'll find those errors. On the other hand, we can say that there are some discrepancies the rules indicate that the expert believes are really not errors. Now the question is: How can we refine and fine-tune the rules to make sure they really don't catch those false-positives? So really the process of fine-tuning the rules is based on comparing the results of the automated assessments with a sampling validation of the data by data experts.

"Data quality scorecards are really the final product of the data quality assessments."

—ARKADY MAYDANCHIK

We've heard a lot about data quality scorecards. In your own words, what really is a data quality scorecard and what role does it play? How is it developed and how does one work in the real world?

Data quality scorecards are really the final product of the data quality assessments. It's a common mistake that a lot of people will think the final product of data quality assessments is a bunch of listings with errors. Say we have a rule that says attributes gender value as M or F and we're going to try to find a list of all records that have other values. Those are the errors. Now if we have 1,000 rules,

we're going to have 1,000 errors. The problem is you're going to end up with listings that have 1,000s or 10,000s of lines, but that doesn't really tell you how good or how bad your data is at the higher level. You really can't translate that into real dollars, you can't really translate it into, "How does it impact our business process? How much does it cost our organization?"

That's where data quality scorecards come into play. It's a hierarchy. On top of it are a few meaningful numbers that we call aggregate scores. Each of them ties the data quality to some specific data use. So, let's say you have a process that uses a certain part of your data, and let's say we ask ourselves, "Given the existing data, how many times, how often is the process going to fail? What is the fraction of the data record that is relevant to this specific process that is correct?" Once we find all the errors form the error reports, we narrow down the targets' subsets that are relevant to this business purpose. Then we calculate the good records among those records that are relevant. So then we can say, "For this specific process, data quality is 95%, and for another process, data quality is 90%, and for a third process or for another use, data quality is 85%." We can differentiate it by business processes by the different groups that use the data and in many other ways. Once we have those aggregate scores, we can basically translate that into the true costs of the business. We can say, "OK, let's see what the significance of any specific type of data error is." Let's say we have a certain kind of data error and it costs us three hours of rework. OK, that translates directly into costs, and now this 95% number can translate it into something meaningful.

> *"We ask ourselves, 'Given the existing data, how many times, how often is the process going to fail?'"*
>
> **—ARKADY MAYDANCHIK**

That's the top level of data quality scorecards. It tells you what the data quality is for different business purposes or different business uses. Of course, below that we have a hierarchy, so we have an ability to drill down and say, "OK, so let's say for this specific business use, data quality is 95%. Well, what's really contributing to that?" Maybe there are 10 different data elements or 100 dif-

ferent data elements that are useful for this purpose. Which of them really contributed to most of the errors? So we can say that data in one of the data elements is 99% and the other one is maybe 89%. That helps us prioritize if we decide to try to improve the data. That helps us prioritize and say, "OK, if we want to improve this number from 95% to 97%, where should we start?" And then the next layer from there is that you can drill down and get into the specific error reports, and that's when you're really looking into any specific data errors and trying to fix them.

So scorecards are a pyramid that provides both the high-level numbers that'll tell you, overall, how good or how bad the data is and how it impacts your business. And on the bottom, the atomic level of data quality that'll tell you if any specific records are good or bad. And then, obviously, there are layers in between that.

What role in a data quality program does the metadata warehouse play and, ultimately, how does it help improve data quality?

The quality of the data is in direct correlation to the quality of the metadata. Let's start with a basic sentence. In a database, we have hundreds of different attributes and each has different meanings, all with different codes that have different meanings. The codes' meanings change over time and can be different for different subsets of the data. Usually, the metadata, which is basically data dictionaries, data models, data catalogues … the metadata, unfortunately, is usually very much out of sync with reality. It's incorrect, it's incomplete, it's obsolete. Organizations usually lack the discipline and tools and staff to ensure that the metadata—the description of the data—is really correct and up-to-date. Now, unfortunately what happens is that the metadata somehow tends to drag the quality of the data down. If the metadata is not accurate, inevitably, over time, the data quality suffers. This happens a lot of times during conversions, any system upgrades and conversion exercises, because usually the conversion mappings are based on the metadata, and if it's wrong, then we end up with wrong data. Data users suffer because people assume that they have something or it means something which it doesn't. And data quality efforts suffer because even when we talk about data quality assessment, we design the rules, and the rules are based on an understanding of what the data is, so if we don't have good metadata, our rules are going to be incorrect and our assessment is going to be incorrect.

So metadata is really very important, and the reality is that we have many different tools to create and collect metadata, we can do data profiling, and we can gather lots of metadata. Also, the assessment itself produces lots of metadata—it produces listings of errors and it produces rules; it's a rules catalog. So there is lots of metadata. You'll find, actually, that the volume of metadata sometimes approaches the volume of the data itself, or at least the complexity, so if you cannot find a way to officially organize your metadata, then you can collect a lot of it, but you cannot use the information and it becomes useless.

So the challenge is a similar challenge to 20 years ago. Our technology made a huge leap forward, and we learned the ability to gather large volumes of data, to organize and process large volumes of data with relational databases and later data warehousing. But the data quality started suffering, since we didn't have the discipline to think about the quality of the data. So now, kind of the same thing with metadata, we have the ability to collect a lot of metadata and to profile data quality assessments, but if we don't plan how to organize all of the metadata we produce, then the same thing happens—the quality of it suffers. If the metadata is no good, the data is no good.

I think a good way to round out our conversation is to get your take on the data quality tools and technologies available on the market. You mentioned that it is still a pretty young market. And what would be your advice to companies just starting out with data quality?

That's a tough question. There are many vendors of data quality and data profiling tools out there. Most of the tools focus on several reasonably narrow areas of data quality management—in those areas, the tools have made tremendous progress and are really very good. For example, if you need a tool for records matching and deduplication, that's something that's been out on the market for seven, eight, 10 years. And there are great tools for that. If you are looking for tools for column profiling, or attribute profiling—things that are going to take attributes one at a time in your database and go through all of the records and give you distributions of values, frequency charts, minimums, maximums, statistical characteristics—there are phenomenal tools. They have come a long way, and they have good interfaces and they do a lot of things.

Unfortunately, at the same time, the tools focus on these few, well-defined, very narrow types of problems, and obviously that was where it

was easiest to start, because everybody had the same problem there. But where we stand now is that everyone is starting to deal with data quality across all kinds of data, and as we want to do a more comprehensive data quality assessment—not just look at names and addresses, not just look at customer data, but look at financial data, HR data, payroll data, insurance claims data—there's really a very big gap between what tools have and what people need. I know that a couple of years ago, in a survey, something to the tune of 80+% of respondents said they were basically custom quoting. And I don't think it has changed much, because as people are looking at data quality problems, they probably find that the tools can only give them a reasonably small mileage. So the bottom line is there is a big need there, on the market, and at this point overall the data quality needs of the customers are lost by the existing vendors. But in certain niches, the tools have made great progress.

Does this mean that companies which want to prove their data quality are going to have to do a lot of this work in-house?

Yes, that's definitely the case. It's kind of a Catch-22. I talk to a lot of vendors, and the reality is there isn't enough demand for some advanced tools for them to make this a serious investment, because the market is still young and most companies are just starting and feeling their way through it. So vendors are not making a leap forward, and they are not creating a tool that could help companies do what they really need to do.

Now part of it, of course, is that up until a few years ago, there wasn't a good place to go to learn how to do it [data quality management]. Right now, between me and a couple of other experts, with books coming out, with all of the training courses that we are teaching at various conferences, there's more and more people. I'd say that just last year I taught more than a thousand people in my classes.

So I'd say we're kind of giving it a new beginning. As more people know how to do it, and what exactly they need to do, they're starting to do it [data quality management]. They are going to vendors with more pointed questions and are asking for more specific features, so I think eventually the market is going to start to pick up. ■

ABOUT THE SPEAKER: **Arkady Maydanchik** is a recognized practitioner, author and educator in the field of data quality and information integration. His data quality management methodology was used to provide data quality services to numerous Fortune 500 companies. Arkady is a frequent speaker at various conferences and seminars, author of the aforementioned *Data Quality Assessment* book, and a contributor to many journals and online publications.

Tips and best practices

This section highlights valuable data quality expert advice and resources from SearchDataManagement.com. Get data quality management tips from some of our most popular experts and read best practices from Q/As and book excerpts.

BY SEARCHDATAMANAGEMENT.COM EDITORIAL TEAM

Like any data management initiative, data quality projects can be tricky for some companies to manage and implement, but with the right knowledge of best practices it can mean fewer headaches for all involved. To help with your data quality projects and processes, we compiled frequently-asked questions answered by experts **David Loshin**, President of Knowledge Integrity, Inc. and **Evan Levy**, Partner and Co-founder of Baseline Consulting. You can read more Q/As at our Ask the Expert section.

QUESTION: What are the best practices in managing data quality, and where can I find a good source of new academic resources in this area?

DAVID LOSHIN: I have a few suggestions. From the academic perspective, I'd check out the MIT Information Quality group and the University of Arkansas Information Quality program, along with the website for the International Association for Information and Data Quality. DAMA has recently released its *DAMA DMBOK (Data Management Book of Knowledge)*, which contains a chapter on data quality.

Most of the larger data quality tool vendors share a significant amount of best practices collateral. Also, check out my comprehensive book on Data Quality, *Enterprise Knowledge Management—The Data Quality Approach.*

QUESTION: What do you think are the most common mistakes that companies make when implementing data quality management programs? We are about to begin an enterprise-wide

data quality management initiative, and I'm wondering whether there are any common pitfalls that we can avoid.

DAVID LOSHIN: Data quality management is not easy, owing to the size and complexity of organizations. There are several common mistakes that companies make during a data quality program implementation. Here are three specific pitfalls that can turn a data quality management project into a nightmare:

① Expecting the silver bullet: Some organizations think they can buy a packaged solution that will address all data quality issues and immediately make them disappear. This unrealistic hope for a "magic tool" is evidenced by how often people acquire a data quality tool as the first step in setting up their data quality program. Buying software before developing a program is indicative of a reactive environment—and the misguided thought that data quality is a technology-driven solution. Too often, senior management gets the idea that you can "fix" noncompliant data instead of eliminating the introduction of bad data in the first place.

How often has your organization bought a tool, only to have it still sitting on the shelf in its shrink-wrap months later? Although data quality tools are critical components of a data quality program, one must first question the motivation for purchasing a tool, then the process itself, and consider the improvement potential in terms of contributing to the effectiveness of the program.

> *"Developing a data quality management program is a strategic undertaking. Its success depends on having both business and technical expertise."*
>
> *—DAVID LOSHIN*

② Not having the right expertise: There is often an expectation that as soon as a data quality program is initiated within an organization, there should be some visible improvement in the data. This is not so. Developing a data quality management program is a strategic undertaking. Its success depends on having both business and technical expertise. This is complicated by the fact that a large part of data quality management, especially at the enterprise level, is advisory.

The close coupling of tools and methods introduces additional complexity to the process. Too often, the data quality manager is viewed as having responsibility for some data quality improvement action without necessarily having either the knowledge or authority to make it happen. The result is an overwhelming feeling that the size of the problem makes its solution unreachable—and consequently the team has no idea where to begin. The mistake occurs in not bringing in the proper expertise to help get the program off the ground.

③ **Not accounting for organizational culture changes:** Even while attempting to improve the quality of data, we often forget that we must work within an organization's existing culture to achieve our improvement goals. No technology in the world will eliminate data quality problems if there is no understanding of how people's behavior allows the introduction of information flaws in the first place.

The evolution of centralized analytical data warehouses provides a good example. Data sets from numerous source systems are extracted, aggregated and transformed in preparation for loading into the warehouse. The data quality problems emerge when the data sets are merged together (perhaps customer names or account numbers are stored in slightly variant forms, data types might not match on similar columns, values are missing, or fields are incomplete). But without the cooperation of upstream systems owners, data warehousing managers are often helpless to control the quality of incoming data. Stricter data quality needs at the data warehouse demand resource allocation by upstream managers. The problem is that their applications may not directly benefit from the desired improvements—and this acts as an effective disincentive for upstream managers to cooperate.

How to avoid data quality management mistakes: Don't despair, though—knowing the common pitfalls of data quality management programs can help you avoid them. Here are some guidelines to keep in mind:

- Exploit the advisory role of data quality teams and use internal procedures to attach responsibility and accountability for data quality improvement to the existing information management authority.

- Don't forget training in the use of policies and procedures—especially in the use of acquired tools.

- Hire professionals with experience in managing data quality projects and programs from the start. They will be able to identify opportu-

nities for tactical successes that together contribute to the strategic success of the program.

- Engage external experts to help jump-start the improvement process. This will reassure your team that your problems are not unique and will allow you to learn from others' best practices.

QUESTION: Where can I find a detailed, free report about the data quality management tools market and a comparative analysis report of various data quality management vendors?

DAVID LOSHIN: If you're looking for unbiased information on data quality management tools, my best recommendation is that you seek out the content available via industry publications and websites, such as The Data Warehousing Institute. Also, certain vendors may make portions of subscription-based analysis (e.g., Gartner Magic Quadrants) available from their websites. For a more comprehensive acquisition effort, I strongly recommend engaging a firm with expertise in the data quality management field to marshal your organization through the requirements analysis/specification, demonstration/proof-of-concept, and assessment phases of the procurement.

QUESTION: How much on average does it cost to clean one customer record? How much should an organization spend on customer data cleansing? Do you know whether there has been any specific reporting or analysis done on this area of customer data quality?

DAVID LOSHIN: The challenge with this question is that underlying its simplicity lie many latent questions whose answers are needed before any kind of customer data cleansing costs analysis can be considered. For example, what data elements constitute the customer record? How many records are there? What are the criteria for declaring a record "clean"? What types of customer data are there? Individuals or organizations? How old are the records? Are they in a single table or scattered across many data assets? What approaches are to be taken for cleansing? There may be studies performed by vendors on the average cost, but I suspect

How much on average does it cost to clean one customer record?

that beneath this question lurk other, more important ones.

To start thinking about the cost of cleansing, consider this example, with residential customer data consisting of first name, last name and telephone number. One can determine whether a single record is "correct" using this algorithm: Call the telephone number, ask to speak with the person whose name shares the record with the telephone number. If the person comes to the phone, ask whether all the values are accurate, and correct those that are not. If there is no one there by that name, the record is incorrect; but, at this point, what can be done to correct it? Either the name is not correct or the number is not correct. The next step in cleaning requires additional information, and if none is available, then the algorithm ends.

Simplistic? Yes. Accurate? Yes. Cost effective? Depends on the number of records, staff members and telephones. Scalable? Not really. There are alternatives, but reliance on different approaches starts to affect those key considerations. Automated solutions may be more scalable, more costly, less accurate, more complex, require more expertise, and so on.

It may be better to challenge the question, then, and turn it into a different sort of beast by suggesting that we answer these questions first and then look at the different alternatives and their corresponding costs:

- What business processes are affected by "unclean" customer data?
- How is "clean" customer data defined?
- What business benefits can be achieved by cleaning customer data?
- What level of precision is necessary for those benefits to be achieved?

The level of effort that is reasonable to spend on customer data cleansing must be less than the value of the accrued business benefits, and this provides an upper limit to what could be budgeted for the process.

QUESTION: Starting off at a high level, what makes a data quality project or program successful? Are there ways a program should or shouldn't be structured to be successful?

EVAN LEVY: One of the biggest challenges when it comes to dealing with data quality is making sure people focus on the fact that data quality is not about data perfection. One of the biggest challenges I find is that people are so focused on "How do I make it better? How do I support my business application users?" They get

very wound up on trying to do what amounts to splitting hairs rather than saying, "Wait a second. If the data is perceived to be bad, what is it that we can do to make it better to support what the business is trying to accomplish?"

So, to answer the question "What makes a data quality project or program successful?" it's about focusing on what we are trying to fix and whether we have bad data differentiating what error detection is.

One good example many people like to use is the address. I can actually determine from someone's street address the city and state they're in if the zip code is missing. With data quality tools and data quality techniques, I can actually distill, interpret or calculate what the zip code field is. However, there are circumstances where because data is incomplete I can't correct it or make it better. So, to sum up the question of how you make the program or project successful, you can't sign up to nirvana, you can't sign up to data perfection because you have to differentiate identifying the error before you can focus on correcting it.

I'd say the other thing to keep in mind when you're dealing with a data quality program is to make absolutely sure you've got someone who's a stakeholder. Whether it's a business user or an application individual, they can give you the guidance and say, "Here's what I need to get out of data quality." There are probably four or five key aspects to data quality that one needs to understand. If I'm determining what the error is, I need to know and agree on the meaning, how I represent it and what the definition of accuracy is. That seems very simple. For example, if I want to define what the color red is [for] my database ... say "Red is a reasonable value for color but I need to make sure that we all agree that that's an accurate value that it's representative in a consistent fashion—R, red or, in fact, the three RGB numeric value of 195 49 28." But one of the benefits of having a business user in place or that applications person is to establish ... what the accurate values might be and whether, in fact, there's enough information to determine what the error value is [and how to correct it]. ■

"The other thing to keep in mind... make absolutely sure you've got someone who's a stakeholder."

—EVAN LEVY

Thirteen causes of enterprise data quality problems

The following is an excerpt from Data Quality Assessment *by Arkady Maydanchik. Reprinted with permission from Technics Publications, LLC. Copyright 2007.* **BY ARKADY MAYDANCHIK**

DATA IS AFFECTED by numerous processes, most of which have an impact on its quality to a certain degree. I had to deal with data quality problems on a daily basis for many years and have seen every imaginable scenario of how data quality deteriorates. Each situation is different, but I eventually came up with a classification shown in **FIGURE 1-1**, page 26. It shows 13 categories of processes that cause the data problems, grouped into three high-level categories.

In this chapter we will systematically discuss the 13 processes presented in **FIGURE 1-1** and explain how and why they negatively affect data quality.

CAUSE NO. 1
INITIAL DATA CONVERSION

Databases rarely begin their life empty. More often, the starting point in their lifecycle is a data conversion from some previously existing data source. And by a cruel twist of fate, it is usually a rather violent beginning. Data conversion usually takes the better half of new system implementation effort and almost never goes smoothly.

CAUSE NO. 2
SYSTEM CONSOLIDATIONS

Database consolidations are the most common occurrence in the information technology landscape. They take place regularly when old systems are phased out or combined. And, of course, they always follow company mergers and acquisitions. Database consolidations after corporate mergers are especially troublesome because they are usually unplanned, must be completed in an unreasonably tight time frame, take

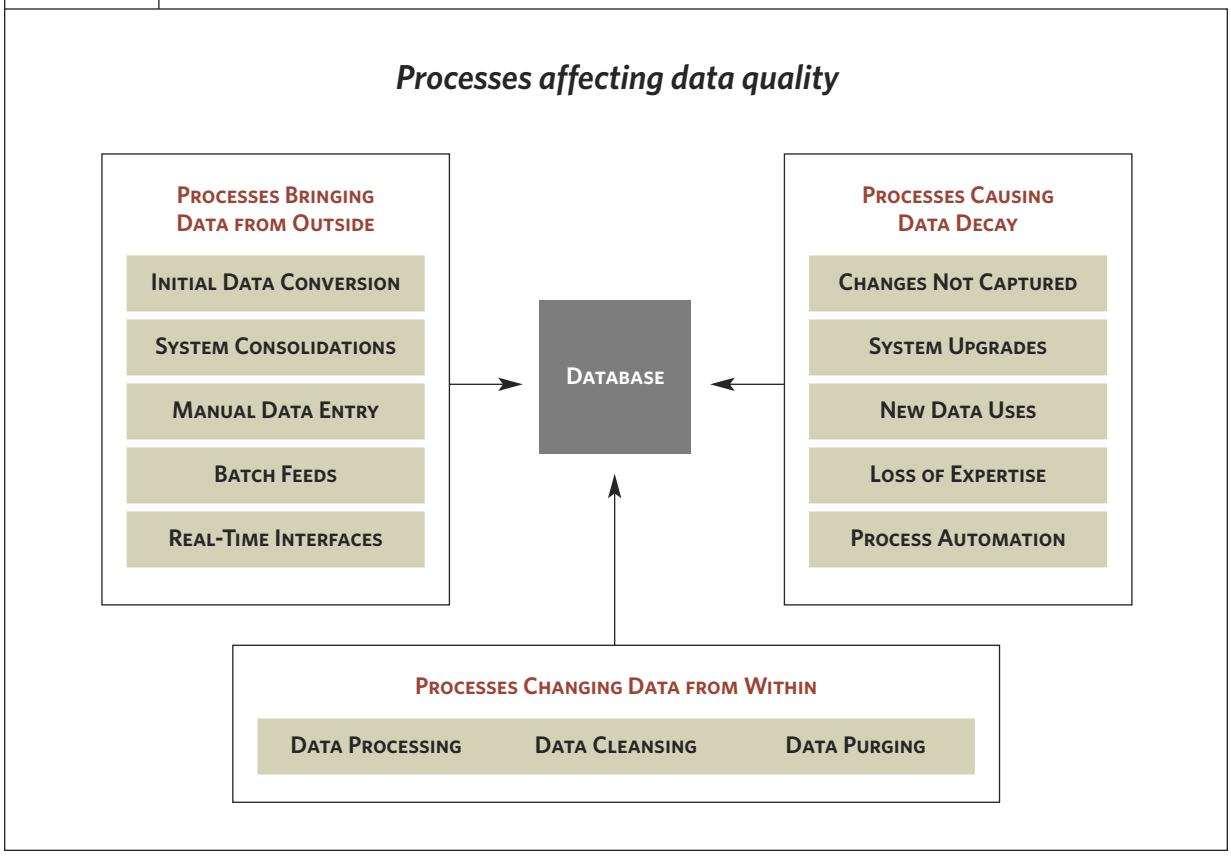
place in the midst of the cultural clash of IT departments, and are accompanied by inevitable loss of expertise when key people leave midway through the project.

CAUSE NO. 3

MANUAL DATA ENTRY

Despite high automation, much data is (and will always be!) typed into the databases by people through various forms and interfaces. The most common source of data inaccuracy is that the person manually entering the data just makes a mistake. To err, after all, is human! People mistype; they choose a wrong entry from the list or enter the right data value in the wrong box. I had, at one time, participated in a data-cleansing project where the analysts were supposed to carefully check the corrections before entering them—and still 3% of the corrections were entered incorrectly. This was in a project where data quality was the primary objective!

FIGURE 1-1



CAUSE NO. 4

BATCH FEEDS

Batch feeds are large, regular data exchange interfaces between systems. The ever more numerous databases in the corporate universe communicate through complex spider webs of batch feeds.

CAUSE NO. 5

REAL-TIME INTERFACES

More and more data is exchanged between the systems through real-time (or near real-time) interfaces. As soon as the data enters one database, it triggers procedures necessary to send transactions to other downstream databases. The advantage is immediate propagation of data to all relevant databases. Data is less likely to be out of sync. You can close your eyes and imagine the millions of little data pieces flying from database to database across vast distances with lightning speed, making our lives easier. You see the triumph of the information age! I see Wile E. Coyote in his endless pursuit of the Road Runner. Going! Going! Gosh!

CAUSE NO. 6

DATA PROCESSING

Data processing is at the heart of all operational systems. It comes in many shapes and forms – from regular transactions triggered by users to end-of-the-year massive calculations and adjustments. In theory, these are repetitive processes that should go "like clockwork." In practice, there is nothing steady in the world of computer software. Programs and underlying data change and evolve, with the result that one morning the proverbial sun rises in the West, or worse yet, does not rise at all.

More and more data is exchanged between the systems through real-time (or near real-time) interfaces.

CAUSE NO. 7

DATA CLEANSING

The data quality topic has caught on in recent years, and more and more companies are attempting to cleanse their data. In the old days, cleansing was done manually and was rather safe. New methodologies have arrived that use automated data cleansing rules to make corrections en masse. These methods are of

great value and I, myself, am an ardent promoter of the rule-driven approach to automated data cleansing. Unfortunately, the risks and complexities of automated data cleansing are rarely well understood.

CAUSE NO. 8

DATA PURGING

Old data is routinely purged from systems to make way for more data. This is normal when a retention limit is satisfied and old data no longer necessary. However, data purging is highly risky for data quality.

CAUSE NO. 9

CHANGES NOT CAPTURED

Data can become obsolete (and thus incorrect) simply because the object it describes has changed. If a caterpillar has turned into a butterfly but is still listed as a caterpillar on the finch's menu, the bird has a right to complain about poor data quality.

CAUSE NO. 10

SYSTEM UPGRADES

Most commercial systems get upgraded every few years. Homegrown software is often upgraded several times a year. While upgrades are not nearly as invasive and painful as system conversions and consolidations, they still often somehow introduce data problems. How can a well tested, better version negatively affect data quality?

CAUSE NO. 11

NEW DATA USES

Remember that data quality is defined as "fitness to the purpose of use." The data may be good enough for one purpose but inadequate for another. Therefore, new data uses often bring about changes in the perceived level of data quality even though the underlying data is the same. For instance, HR systems may not care too much to differentiate medical and personal leave of absence—a medical leave coded as a personal leave is not an error for most HR purposes. But start using it to determine eligibility for employee benefits, and such minute details become important. Now, a medical

The risks and complexities of automated data cleansing are rarely well understood.

leave entered as a personal leave is plain wrong.

CAUSE NO. 12

LOSS OF EXPERTISE

In almost every data quality project on which I worked, there is a Dick or Jane or Nancy whose data expertise is unparalleled. Dick was with the department for the last 35 years and is the only person who really understands why for some employees date of hire is stored in the date of birth field, while for others it must be adjusted by exactly 17 days. Jane still remembers times when she did calculations by hand and entered the results into the system that was shut down in 1985, even though she still sometimes accesses the old data when in doubt. When Nancy decided to retire, she was offered hourly work from home at double her salary. Those are true stories.

CAUSE NO. 13

PROCESS AUTOMATION

With the progress of information technology, more and more tasks are automated. It starts from replacement of data entry forms with system interfaces and extends to every layer of our life. Computer programs process and ship orders, calculate insurance premiums, and even send spam—all with no need for human intervention. Where in the past a pair (or several pairs) of human eyes with the full power of trained intellect protected the unsuspecting customers, now we are fully exposed to a computer's ability to do things that are wrong and not even feel sorry. ■

Arkady Maydanchik is a recognized practitioner, author, and educator in the field of data quality and information integration. His data quality management methodology was used to provide data quality services to numerous Fortune 500 companies. Arkady is a frequent speaker at various conferences and seminars, an author, and a contributor to many journals and on-line publications.

Data quality assurance

The following is an excerpt from Data Quality: The Accuracy Dimension *by Jack E. Olson. Reprinted with permission from Morgan Kaufmann, a division of Elsevier. Copyright 2003.*

BY JACK E. OLSON

GOALS OF A DATA QUALITY ASSURANCE PROGRAM

A data quality assurance program is an explicit combination of organization, methodologies and activities that exist for the purpose of reaching and maintaining high levels of data quality. The term assurance puts it in the same category as other functions corporations are used to funding and maintaining. Quality assurance, quality control, inspection, and audit are terms applied to other activities that exist for the purpose of maintaining some aspect of the corporation's activities or products at a high level of excellence. Data quality assurance should take place alongside these others, with the same expectations.

Just as we demand high quality in our manufactured products, financial reports, information systems infrastructure, and other aspects of our business, we should demand it from our data.

The goal of a data quality assurance program is to reach and maintain high levels of data accuracy within the critical data stores of the corporation. It must encompass all existing, important databases and, crucially, be a part of every project that creates new data stores or that migrates, replicates or integrates existing data stores. It must address not only the accuracy of data when initially collected but accuracy decay, accurate access and transformation of that data, and accurate interpretation of the data for users. Its mission is threefold: improve, prevent, monitor.

Improvement assumes that the current state of data quality is not where you want it to be. Much of the work is to investigate current databases and information processes to find and fix existing problems. This

effort alone can take several years for a corporation that has not been investing in data quality assurance.

Prevention means that the group should help development and user departments in building data checkers, better data capture processes, better screen designs, and better policies to prevent data quality problems from being introduced into information systems. The data quality assurance team should engage with projects that build new systems, merge systems, extract data from new applications, and build integration transaction systems over older systems to ensure that good data is not turned into bad data and that the best practices available are used in designing human interfaces.

Creating a data quality assurance program and determining how resources are to be applied needs to be done with careful thought.

Monitoring means that changes brought about through data quality assurance activities need to be monitored to determine whether they are effective. Monitoring also includes periodic auditing of databases to ensure that new problems are not appearing.

STRUCTURE OF A DATA QUALITY ASSURANCE PROGRAM

Creating a data quality assurance program and determining how resources are to be applied needs to be done with careful thought. The first decision is how to organize the group. The activities of the group need to be spelled out. Properly skilled staff members must be assigned. They then need to be equipped with adequate tools and training.

DATA QUALITY ASSURANCE DEPARTMENT

There should be a data quality assurance department. This should be organized so that the members are fully dedicated to the task of improving and maintaining higher levels of data quality. It should not have members who are part-time. Staff members assigned to this function need to become experts in the concepts and tools used to identify and correct quality problems. This will make them a unique discipline within the

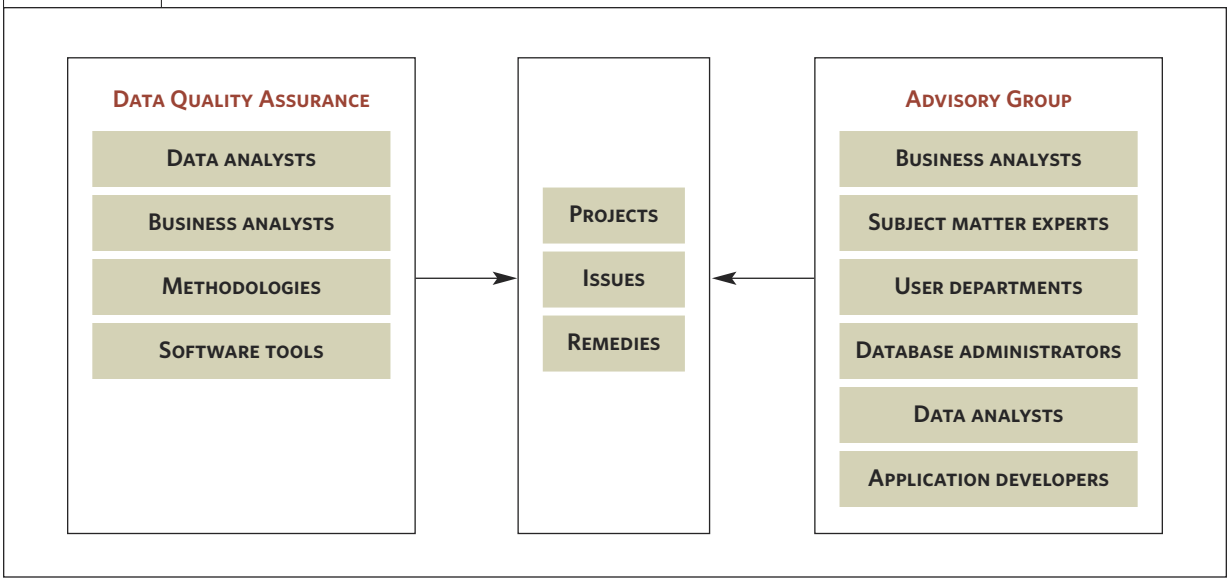
corporation. **FIGURE 4.1** is a relational chart of the components of a data quality assurance group.

The group needs to have members who are expert data analysts. Analyzing data is an important function of the group. Schooling in database architecture and analytical techniques is a must to get the maximum value from these activities. It should also have staff members who are experienced business analysts. So much of what we call quality deals with user requirements and business interpretation of data that this side of the data cannot be ignored.

The data quality assurance group needs to work with many other people in the corporation. It needs to interact with all of the data management professionals, such as database administrators, data architects, repository owners, application developers, and system designers. They also need to spend a great deal of time with key members of the user community, such as business analysts, managers of departments, and Web designers. This means that they need to have excellent working relationships with their customers.

One way to achieve a high level of cooperation is to have an advisory group that meets periodically to help establish priorities, schedules and interactions with the various groups. This group should have membership from all of the relevant organizations. It should build and maintain an inventory of quality assurance projects

FIGURE 4.1



that are worth undertaking, keep this list prioritized, and assign work from it. The advisory group can be very helpful in assessing the impact of quality problems as well as the impact of corrective measures that are subsequently implemented.

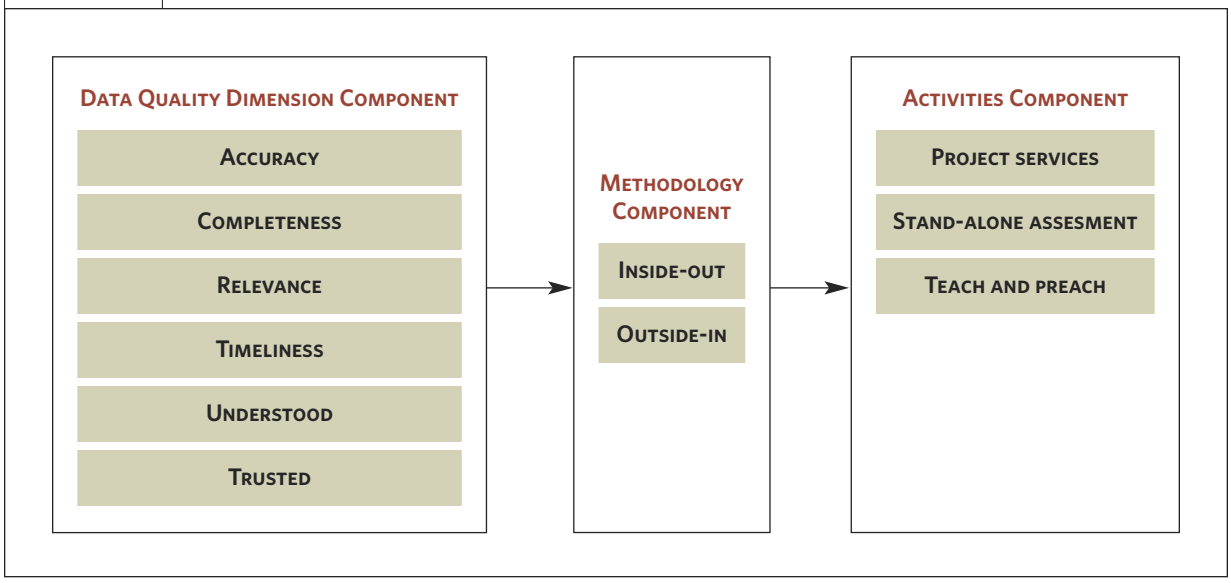
DATA QUALITY ASSURANCE METHODS

FIGURE 4.2 shows three components a data quality assurance program can build around. The first component is the quality dimensions that need to be addressed. The second is the methodology for executing activities, and the last is the three ways the group can get involved in activities.

The figure highlights the top line of each component to show where a concentration on data accuracy lies. Data accuracy is clearly the most important dimension of quality. The best way to address accuracy is through an inside-out methodology, discussed later in the book. This methodology depends heavily on analysis of data through a process called data profiling. The last part of this book is devoted to explaining data profiling. Improving accuracy can be done through any of the activities shown. However, the one that will return the most benefit is generally the one shown: project services.

Any data quality assurance function needs to address all of the dimensions of quality. The first two, data accuracy and completeness,

FIGURE 4.2

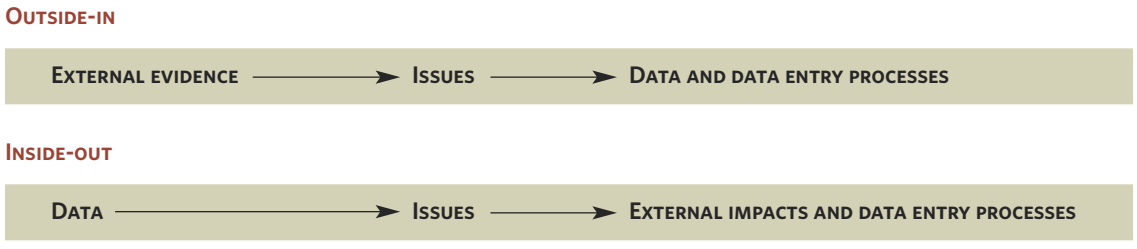


focus on data stored in corporate databases. The other dimensions focus on the user community and how they interpret and use data.

Issues are actionable items: They result in activities that change the data quality of one or more databases.

The methods for addressing data quality vary, as shown in **FIGURE 4.3**. Both of these methodologies have a goal of identifying data quality issues. An issue is a problem that has surfaced, that is clearly defined, and that either is costing the corporation something valuable (such as money, time or customers) or has the potential to cost the corporation something valuable. Issues are actionable items: They result in activities that change the data quality of one or more databases. Once identified, issues are managed through an issues management process to determine value, remedies, resolution, and monitoring of results. The process of issue management is discussed more fully in the next chapter. ■

FIGURE 4.3



Data quality: Why management should care about bad data

The following is an excerpt from Data Quality: The Field Guide *by Tom Redman. Reprinted with permission from Digital Press. Copyright 2001.* **BY TOM REDMAN**

SOME THINK THAT no two words can cause a CEO's eyes to glaze over faster than "data quality" (and this applies to heads of government agencies, leaders of nonprofit organizations, etc.). "Data," aren't they the boring bits and bytes buried in our computer systems? And "quality," isn't that the implication that our people aren't working hard enough?

Besides, people have real work to do, customers to satisfy, production schedules to meet, decisions to make, strategies to map out, a demanding board to satisfy. Who wants to worry about those bits and bytes when no one is complaining?

But CEOs are (or should be) passionately interested in data quality, and for a wide variety of reasons.

First, bad data can earn the CEO and his or her organization a place in the national news—and who needs that? The bombing of the Chinese Embassy is the most publicized recent example. But it happens more frequently than one might think.

Fortunately, most cases of bad data do not land the organization or its leader on the front page. Unfortunately, poor-quality data seems to be the norm. As CEOs know, the costs of poor-quality data are enormous. Some costs, such as added expense and lost customers, are relatively easy to spot, if the organization looks. We suggest (based on a small number of careful, but proprietary, studies), as a working figure, that these costs are roughly 10% of revenue for a typical organization. To date, no one, in hundreds of discussions, has suggested that this number is "way too high." CEOs naturally want to return these monies to the bottom line. ■