Norfolk Southern Travels the BI Track
Hugh J. Watson, Barbara H. Wixom, and Linda Richardson

Pervasive Business Intelligence and the Realities of Excel
Russell Tront and Suzanne Hoffman

Agile Project Leaders Redefine IT (Innovative Thinkers)
Valena Hernandez

BI Experts’ Perspective: Moving to Real-Time BI
Sami Akbay, Harriet Fryman, David Stodder, and James Taylor

Harnessing the Power of Social Media Data with Master Data Management
Ravi Shankar

Four Strategies for Increasing Business Intelligence Usage
Khwaja Saad Waheed

BI Case Study: Apparel Company Melds Fashion, Mobile BI
Linda L. Briggs

Making Your Organization Care about Data Quality
Mario Cantin

BI Bookshelf: Enabling BI Agility with Data Virtualization
James E. Powell
BI Training Solutions:
As Close as Your Conference Room

We know you can’t always send people to training, especially in today’s economy. So TDWI Onsite Education brings the training to you. The same great instructors, the same great BI/DW education as a TDWI event—brought to your own conference room at an affordable rate.

It’s just that easy. Your location, our instructors, your team.

Contact Yvonne Baho at 978.582.7105 or ybaho@tdwi.org for more information.

tdwi.org/onsite
3 From the Editor

4 Norfolk Southern Travels the BI Track
Hugh J. Watson, Barbara H. Wixom, and Linda Richardson

8 Pervasive Business Intelligence and the Realities of Excel
Russell Tront and Suzanne Hoffman

15 Agile Project Leaders Redefine IT (Innovative Thinkers)
Valena Hernandez

19 BI Experts’ Perspective: Moving to Real-Time BI
Sami Akbay, Harriet Fryman, David Stodder, and James Taylor

27 Harnessing the Power of Social Media Data with Master Data Management
Ravi Shankar

34 Four Strategies for Increasing Business Intelligence Usage
Khwaaja Saad Waheed

39 BI Case Study: Apparel Company App Melds Fashion, Mobile BI
Linda L. Briggs

42 Instructions for Authors

43 Making Your Organization Care about Data Quality
Mario Cantin

53 BI Bookshelf: Enabling BI Agility with Data Virtualization
James E. Powell

56 BI StatShots
From the Editor

Ever since my local PBS affiliate began airing the British TV series *New Tricks*, I’ve been hooked. The program follows a group of three retired detectives who, together with a fourth detective currently on the force, tackle cold or unusual cases. It’s *CSI* meets the over-60 set.

It’s fascinating to watch the melding of good, old-fashioned police work with newer tools and technology—valuable experience is combined with out-of-the-box thinking to solve cases that occurred when DNA tests were unheard of. In much the same way, this issue of the *Business Intelligence Journal* looks at how newer approaches are tackling old business intelligence (BI) problems.

The *Journal*’s senior editor Hugh Watson and associate editor Barbara Wixom examine Norfolk Southern’s path along the BI maturity model and reflect on how new environmental changes drive the need for new business strategies—and thus new BI capabilities (read: new tricks).

Mario Cantin examines another new trick: linking the quality of data to project benefits and developing practical ways to ensure your organization cares about the quality of its data. Also on the forefront: data virtualization, which is the subject of our book review.

One old and ever-popular BI tool is Excel. Russell Tront and Suzanne Hoffman explore how Excel usage is changing, especially in BI. They provide tips and best practices (that is, new tricks) for using Excel to make BI more pervasive. Khwaja Saad Waheed provides strategies to make BI more pervasive, too. Some are old (I first saw drill-down capabilities in the personal financial manager Quicken many, many years ago); others are new (such as action-integrated BI).

Ravi Shankar examines how an old tool—one-to-one marketing—can use the new data-rich sources of social media to enhance BI results. Valena Hernandez takes us inside an exercise at a TDWI BI Executive Summit to show how old habits must give way to new thinking in order to be agile.

Our four experts tackle the struggle enterprises have long faced when moving to real-time BI. They discuss the impacts to BI staff, the new technologies (and tricks) needed, and ways to engage employees to overcome resistance to change. Our case study looks at something that must always be new—fashion—and shows how one firm incorporated a new trick—mobile access—to change how BI was used.

We’re always interested in your comments about our publication and the articles you find most valuable for helping your business solve old challenges with new thinking, approaches, and tools. Please send your comments to jpowell@tdwi.org.
Business intelligence (BI) managers must be both reactive and proactive. They must be reactive because a change in the business environment can lead to changes in the business strategy, which in turn leads to new BI requirements. They must be proactive because the demand for new BI capabilities requires having the people, processes, and technology in place to meet the demand.

For the environment-strategy-BI link to work well, companies must have people and processes in place (e.g., steering committees) to ensure that new business strategies and plans are communicated and supported by BI. To ensure that BI can support the new demands, an appropriate BI infrastructure must be in place. In most cases, the development of this BI infrastructure is a long-term, evolutionary process, which can be described by a BI maturity model. For example, TDWI’s maturity model developed by Wayne Eckerson has prenatal, infant, child, teenager, adult, and sage stages (the model is available at tdwi.org/bimaturityassessment; short registration may be required). Firms must be mature along multiple dimensions—scope, sponsorship, funding, delivery or reporting, development, data, architecture, and value—all of which contribute to the firm’s BI capabilities.

Norfolk Southern Corporation (NS) provides an excellent example of an organization that has responded to changing environmental conditions with new business strategies and greater BI maturity. There are lessons to be learned from the company’s experiences, and I’ve invited Barbara Wixom from the University of Virginia and Linda Richardson from Norfolk Southern to help describe them. A more extensive version of the Norfolk
Southern BI story is available in *MIS Quarterly Executive* (Wixom, 2011).

**About Norfolk Southern Corporation**

Norfolk Southern is one of the four largest railroad companies in the U.S. The railroad has approximately 20,000 route miles in 22 eastern states and the District of Columbia, serves all major eastern ports, and connects with rail partners in the West and Canada, linking customers to markets around the world. NS provides comprehensive logistics services and offers the most extensive intermodal (transportation of freight using multiple modes—rail, ship, or truck) network in the East.

**Environmental Changes Require New Business Strategies and BI Capabilities**

For decades, NS operated in a predictable, regulated environment, and the company’s business strategy focused on efficiently transporting freight from one point to another. Unfortunately, transactional systems could not be used for any significant reporting and analytical purposes because queries could possibly degrade system performance. When employees needed a report, they had to submit a request to IT. NS began its BI journey in the infant stage.

This strategy began to change after deregulation, and by the mid-1990s the marketing and cost departments championed the development of a Teradata data warehouse to allow managers to access the vast amount of data that the transactional systems were producing. Over the next few years, NS moved into the teenager stage.

The railroads began to compete in new ways. In response, NS adopted a growth strategy and added more than 6,000 miles of track through acquisition. The company also became much more service oriented.

NS created the Thoroughbred Operating Plan (TOP) in 2002 to support the company’s competitive position. The plan called for investment in new transactional systems and a heavy dose of BI, including the use of optimization algorithms, improved performance monitoring systems, and customer access to current shipment data. NS was moving into the adult BI stage.

In recent years, NS developed a strategy called Track 2012 that leverages BI across the enterprise. This strategy puts customer service at the forefront. It also focuses on the management of the company’s cost structure, which has three primary drivers—fuel, the productivity of locomotives and cars, and workforce productivity. BI applications are in place to address all of the business strategy components and have moved NS solidly into the adult stage of BI.

**Lessons Learned along the Way**

Along the path to becoming a mature BI organization, Norfolk Southern has progressed technically, managerially, and organizationally.

**Lesson #1: Make users self-sufficient.**

In mature BI organizations, BI is pervasive and available to customers, suppliers, and employees throughout the organization. A potential obstacle to pervasive BI is the cost and difficulty of supporting a large user base. At NS, the BI team and sponsors have proactively worked to make internal and external users self-sufficient. This goal is accomplished with portals, point-and-click interfaces for querying and dashboards for monitoring, and wizards for building new reports. Major BI applications at NS require little training.

**Lesson #2: Establish business-run governance.**

Over time, a variety of BI sponsors have emerged throughout the organization. In 2007, NS formally organized these sponsors into the Teradata Users Group (TUG). The TUG brings together 24 business managers...
and three representatives from IT who meet for two hours every month. The TUG chair comes from the business, and each business unit (including IT) has one voting right. The TUG creates special interest groups (SIGs) to focus on special issues, such as to drive metadata initiatives, investigate data quality issues, and select a next-generation BI tool for the company.

The TUG is a vehicle for sharing information and spreading BI best practices. It allows business groups to voice opinions about BI, alter BI priorities, and provide input into BI decisions. Attendance at TUG meetings is taken very seriously, as TUG members cannot delegate their spots to others.

The TUG members have learned a great deal about BI at NS through information shared by their peers. At each meeting, a subject-area expert describes existing data or applications to the TUG. These presentations raise awareness of the data that is available and suggest possible new applications.

Lesson #3: Develop “hybrid” employees with both business and technical skills.
In many organizations, there is a chasm between the IT people who build applications and the people who use them. NS addresses this chasm with hybrid employees who have both domain knowledge and technical skills. The company creates these “hybrids” by moving IT people into the business groups, transferring people from the business groups into IT, hiring people with both technical and business skills, and creating BI groups within business units.

For example, when one employee first came to work for NS, he was in the IT group and was assigned to work with financial systems. Over time, the financial systems became highly automated, and to manage and operate them, it was necessary to understand both accounting and technology. This employee is now officially a part of the accounting department, where his strong technical skills are highly useful.

Lesson #4: Have scalable BI technology and get the most out of it.
There is a potential downside to high levels of BI maturity—ever-growing volumes of data, a larger user base, and more complex queries—all of which can strain the underlying technical platform and result in costly upgrades. Consequently, companies should invest in technology that scales over time and tune it regularly.

NS selected a platform that has scaled to meet its needs, but upgrades are costly and NS has focused on getting as much performance as possible out of what is already in place. It initiated a project called PET, or “performance enhancements on Teradata.” Over time, the PET team has become a group that ensures that NS is a responsible owner of data warehouse technology.

Once a month, database administrators, production support, new development, and BI software groups meet and review reports on processor and system utilization, which are broken down by user, applications, batch job, and so on. The team investigates what consumes considerable resources and then crafts strategies that increase efficiency. These improvements can mean adding an index to a table, turning off a process that doesn’t need to be running, reworking a SQL query, better training users who perform ad hoc queries, or moving a process to occur at an off time. Over a period of 18 months, the
PET group kept CPU usage flat, which translates into significant technology cost savings.

Lesson #5: View data as an enterprisewide asset.
In a mature BI organization, the data warehouse is enterprisewide in scope, containing subject areas that span the organization. This data supports existing applications and makes it easy to quickly develop new ones.

NS has followed several key practices in developing its warehouse. The same contractor has modeled the data since the inception of the warehouse. This person began by modeling the data with an enterprise approach and was able to create standards and maintain consistency as new subject areas were added. In addition, the TUG drives special initiatives to develop data as an enterprise asset. For example, the TUG formed SIGs to work on metadata and data quality issues.

NS has an “open-door” policy whereby user groups are encouraged to share data. This sharing has led to new, interesting uses of data. For example, the human resources department combined employee demographic data (e.g., ZIP codes) with geospatial data traditionally used by the engineering group to visually map out employee population density, making it much easier to locate offices for services for NS employees.

Conclusion
It is common to talk about an organization being in a particular state of maturity, such as an adult stage. It is important to recognize, however, that there are multiple dimensions (or characteristics) that work together to establish an organization’s overall or composite stage. Each of these dimensions progresses through its own set of maturity stages. Although all of the stages for the dimensions are seldom exactly in sync, they should be at approximately the same stage of evolution or problems will occur.

NS has done a great job of advancing along multiple dimensions. The company has implemented practices to make BI more pervasive (the scope dimension); created business-run governance structures (the sponsorship dimension); invested in scalable technology (the architecture dimension); and treats data as an enterprisewide asset (the data dimension). The end result is great business value.

Reference
Pervasive Business Intelligence and the Realities of Excel

Russell Tront and Suzanne Hoffman

Abstract
For top performance, enterprises need to take advantage of business intelligence (BI) at every level. Microsoft Excel is a predominant tool for ad hoc BI. Most BI vendors have an Excel BI component (with varying degrees of functionality) because BI data will eventually end up in Excel. In this article, we discuss the evolving nature of BI usage, present survey results on the use of Excel in BI, and provide tips and best practices for deploying Excel and non-Excel BI while enabling pervasive BI.

The Evolving Nature of Pervasive Business Intelligence
An emerging core competency of every thriving business is the ability to deliver real-time analytical intelligence throughout the company. More important than the selection of the “right” presentation solution is the delivery of the right data, with the appropriate level of granularity and ad hoc interactivity, to all levels of decision makers in the enterprise. This concept of pervasive BI empowers employees to understand and meet their individual goals and corporate objectives from “the shop floor to the top floor.”

BI—and the process by which decisions are made in enterprises of all sizes and in all public and private sectors—enables personnel to contribute to the overall health of the organization, regardless of their positions. Enabling pervasive BI helps to buoy morale and better align departments and decision making with more global objectives. Instead of making 20 or 30 decisions that affect everyone, an additional, far larger number of microdecisions can be made that are course corrections and fine-tuning of business and operational strategies. As a result, a company can compete on the strength and the quality of its decision-making ability, not just on the
goods and services it provides. Changing the culture and capability of the organization so that every decision has value, even at a very tactical level, delivers on the promise of pervasive BI as a competitive advantage.

The closer the decision is to the generation of the data, the more informed the decision will be. Traditional deployment of BI systems focused on the user interface and tools that were provided. Organizations made BI technology decisions by focusing on a small group of core users who were typically in the administration area of the organization and did not represent all decision makers. Attempts to cascade the use of BI tools further into departments—the “democratization of BI”—were not readily embraced. Tools were seen as frustrating and data requests created backlogs of IT requests. Executives and managers maintained that workers did not need BI tools, but in reality, most of these workers already had a tool at their fingertips that they knew well and could use effectively with minimal training—Microsoft Excel.

How much of any BI technology environment is ever fully utilized? Industry pundits share statistics that an occasional BI consumer knows only 10 to 20 percent of any technology features. They are consumers of the data, but not analytic users of the tool. If the goal is pervasive BI—an environment in which 40 to 60 percent of a company’s employees make business decisions based on the data generated by the company’s systems—then the BI tool must not be the limiting factor.

Basic mathematics classes as early as high school teach statistics and comparative analysis to students. Data is represented in a cross-tab or a spreadsheet to help students visualize the relationships. Typically this data is manually entered and algorithmically represented in relationship to other variables and data. Fast-forward 10 years, and these same students are the twenty-something employees that we need to teach to understand the decision-making process in any business or service sector. Excel is a valid choice as either a primary or secondary BI tool if its use is implemented correctly—for example, with hot, refreshable connections to the data of record.

BI technology vendors have been consistent in their antipathy to Excel. We are not speaking of Microsoft BI, but more generically about Excel or any other cross-tab program. Detractors complain about “spreadmarts,” “Excel hell,” and other common user-created environments where large numbers of spreadsheets become the source for unattributable data (data that was manually entered or copied from another system and no longer has lineage to the data store of record).

BI as a technology-vendor segment is focused on end-user tools. However, BI as an industry initiative is really about the data. If we transition away from the “what” to the “why,” we must look at the data available to the BI user. Data is both fabricated and captured—internally and externally. The cleansing, structuring, reformatting, and mashing during analysis preparation is the lifeblood of any organization and the source of its competitive intelligence. Decision making is only as good as the data upon which decisions are made. We are reminded of W. Edwards Deming’s quip, “In God we trust; all others must bring data.” If the data is consistent and the tools are always refreshed and tethered to the data store of record, then any good, easy-to-use tool will be generally satisfactory. The key to the decision is the data.

Enabling pervasive BI helps to buoy morale and better align departments and decision making with more global objectives.

Although the concept of pervasive BI has been a goal of many enterprises, it has been out of the reach of most. IT has tried to consolidate the tools and technologies to a single vendor. The typical approach is for the vendor to deliver a range of services to users, who are characterized by their data consumption and analytical sophistication. We see a breadth of functionality offered by many BI
vendors to accommodate a range of users. For example, vendors have been quick to add mobility to their stacks to accommodate the disconnected user and global executive. Constant corporate evolution and a fluctuating economy have made adoption difficult as decisions need to be made in real time and enterprises don’t have the luxury of a low learning curve or new purchase.

The promise of pervasive BI also has some interesting implications about the alignment of the types of decisions that can be made in an organization. Increased interest in employee performance management tied with incentive pay can be directly attributable to adoption of BI as a measure of personal success. Corporate reorganizations, mergers, and acquisitions have less impact if the data is immediately accessible to everyone. These areas are just a few of the benefits of a pervasive BI–enabled environment.

When we consider the evolution to “big data” and data warehouse stores that exceed multi-petabytes, disparate systems can typically be commingled and represented in the correct grain of detail for any decision being made. As companies evolve and rethink their traditional architectures, databases that exceed the capabilities of mundane data storage are fast becoming a reality. Mixed-use storage, once a struggle, is now common. Databases with online analytical processing (OLAP) components either inside or as a virtual layer are the mainstay of data warehousing. Columnar databases, analytic databases, and algorithmic databases all support both SQL and multidimensional expression (MDX) query languages when positioned for data warehousing. Midtier servers supporting single BI tools as a semantic layer may become obsolete. This transition in the industry is enabling Excel and other tools to become more widespread and acceptable as complements to existing BI deployments.

The roles of users making business decisions will cross management boundaries to engineers, marketing assistants, call center representatives, sales personnel, software developers, and others. It stands to reason that all of these individuals, when asked to make decisions that affect their jobs and their departments, should be able to access the appropriate data and use a BI tool that is easily understood, yet powerful enough to do the types of analyses needed. The best training should focus on the data and the types of analytics that should be performed. The best tool for the job is the one best known at the time the analysis must be performed.

Consumers of data need an analysis tool commensurate with the complexity of their role, the data, and the expectations of their analysis. The data should be equally accessible by the tool of their choice regardless of where the data resides.

As the data warehouse market evolves, database vendors are enabling Excel access using MDX to accommodate the increasing use of this ubiquitous spreadsheet tool. All OLAP vendors export analyses to Excel, and most software packages have Excel add-ins. We could not find any empirical data about the use and acceptance (or nonacceptance) of Excel by corporations and IT departments, however. If vendors are being asked to accommodate their users’ requests for Excel support, where is the data supporting this acceptance of the much-maligned spreadsheet?

How Do Excel and BI Integrate in an Organization?
Unable to find supporting data, we conducted an industrywide survey on the use of Excel as an ancillary tool to BI. The goal was to better understand how companies currently use Excel in a BI environment and whether it could be a critical component of pervasive BI deployments.

As companies evolve and rethink their traditional architectures, databases that exceed the capabilities of mundane data storage are fast becoming a reality.
More than half of our respondents were business analysts—the crossover role between IT and a typical business user. Another 21 percent were pure business users and 19 percent were from IT. This mix provided feedback from both sophisticated users and the users who relied more on IT to deliver their reports or BI data. Respondents came from a variety of industries: education, finance, manufacturing, telecommunications, construction, retail, healthcare, and government. The size of the organizations varied, with 30 percent being under $10 million in revenue and 30 percent earning over $1 billion. The demographics of respondents were equally broad based. From a role perspective within the respective organizations, we had a good cross section of management and executive representation, but a strong nonmanagerial contingent of 48 percent.

As a group, the majority of the respondents are longtime Excel users. Over half (54 percent) have been using Excel for more than 15 years. In contrast, the average time spent with their current BI tool was four years. Over one-quarter (28 percent) said that their company had abandoned one or more BI tools in the past 10 years. Although Excel certainly has staying power as a desktop tool, it is also capable of being used responsibly as an analytic tool. If Excel is deployed in a secure manner (so that a shared spreadsheet, when refreshed, reconnects to the original data source), nearly nine in 10 respondents (88 percent) said that sharing spreadsheets would no longer be alarming.

Tethering Excel with a hot connection to the data store will also help prevent basing decisions on inaccurate or misleadingly old data, which 58 percent of our respondents reported has happened. When asked about data lineage and the source data for an Excel spreadsheet, 46 percent said they were consistently skeptical of the origin of the data, but not skeptical enough to stop using Excel.

Non-IT respondents reported using 22 different BI tools in all. Their degree of comfort with their tools is high (at 69 percent). The one tool they prefer over all others is Excel; more than 50 percent said they routinely export data to Excel and more than 92 percent said the data eventually makes its way into Excel at all levels of the corporate hierarchy.

Clearly there is a missed opportunity to incorporate the innate BI capabilities of Excel into the BI culture and ensure that it has direct connectivity to the BI data, thereby maintaining data lineage and integrity. On a scale of one to five (with five being an expert ranking), respondents ranked themselves as experienced (a four) with the features of Excel that would be used in a BI environment, including data loading, data manipulation, PivotTables, graph and chart creation, and working with formulas and conditional logic.

Excel 2007 is the predominant version of Excel being used. SharePoint Excel Services and the lack of SharePoint adoption was a bit surprising. Only 22 percent said that their organization had an ongoing initiative to explore using SharePoint to host data-connected spreadsheets that are viewable and refreshable by a variety of personnel.

There is a reluctant user community that did not join in the survey. Survey questions addressed specifically to IT respondents revealed that 56 percent of their users ask for custom BI reports more than once a week. These custom reports are a huge drain on time and resources and defeat the movement toward self-service BI. This response was surprising, as was the frequency of refresh required: 42 percent said the data was refreshed daily and 15 percent said it was refreshed in real time. With 90 percent of the non-IT users reporting that they build their own reports, the group that is plaguing IT obviously did not participate in the survey. Our survey drew a high response rate from business analysts—they know their data, need it to be timely, and are performing sophisticated analyses. They have identified themselves as our BI gurus, and they still use Excel.

When asked about data governance, only 12 percent of the IT respondents said they monitor the use of Excel as a BI tool. As a group, IT indicated they would like more control and oversight, but only 33 percent said they were worried about security of the data. The most commonly enabled security feature reported is secure sign-on to their
BI data. IT users of Excel seemed much more inclined than end users to experiment with the new features of Excel 2010, which bodes well for adoption of the software’s more advanced BI functionality.

Our experience with companies that embrace Excel as a complement to their existing BI implementation is that they have done so to satisfy a group of users who were intimately familiar with the data but less familiar with available BI tools. Users turn to Excel because it is the best-known environment; it is easy and flexible to use and does not require IT support. Technology has advanced in a way that allows Excel to be integrated into an enterprise BI environment and enable a class of users that may have otherwise been overlooked as potential users of BI data. For example, we know of a major bank that has a large group of risk managers who typically use SAS for their mathematical algorithms and statistics, but use Excel directly connected to Oracle OLAP when performing comparative analytics and portfolio analysis.

At a large automobile manufacturer, a group of engineers (a rather atypical BI user group) much prefer Excel for quality assurance and Six Sigma analysis. Retail users, who routinely use a formal BI tool and process, will take an extract of data into Excel to perform variance analysis that is not part of their dashboard. These instances exemplify self-service, ad hoc BI where the user is enabled with data, not disabled with technology.

**Best Practices for Managing Pervasive BI**

Having discovered how pervasively Excel is used in BI, we compiled the following four tips for applying Excel in a variety of BI environments. Many of these practices are very general and also apply to BI solutions such as Kognitio Pablo, Teradata Business Intelligence Optimizer, SAP BusinessObjects, and Simba’s MDX Provider for Oracle OLAP.

**Tip #1: Conduct user training.**

Train users to not copy or perform single-instance data import into Excel and base analysis on this quickly stale data. Rather, mandate that users connect externally to the data source of record from Excel, and train them how to connect to and easily refresh from their OLAP source data. Instruct users about storing their data source user name and password in a Windows data source name (DSN) or in an Office data connection (.odc) file, if your organization allows it. Alternately, investigate single sign-on (SSO) or Windows authentication options, if your data source can use them. Users will be more likely to keep their data up to date when they use a hot connection to your OLAP data source—and thus avoid spreadsmarts.

IT users of Excel seemed much more inclined than end users to experiment with the new features of Excel 2010, which bodes well for adoption of the software’s more advanced BI functionality.

**Tip #2: Make file sharing easier.**

Make it easier for users to share spreadsheets that have external data connections. Teach users how to create and pass file DSNs or how to mount an .odc file on your network or SharePoint server, and have all users connect using it.

**Tip #3: Use advanced PivotTable features.**

One reason people sometimes use disconnected data is because they want a table laid out differently from a standard Excel PivotTable. An Excel PivotTable is an indivisible object; users can’t just insert a useful row or column into the middle of it. Similarly, sometimes users don’t know how to get rid of annoying hierarchy-level subtotals. To get only the format desired, users may copy and paste the PivotTable values to another worksheet, and insert local rows/columns and calculations from there. Often this data quickly becomes stale as it is mailed or shared within an organization. Here are two suggestions to overcome this problem:
1. Manipulate totals. Teach your users how to show or hide totals or levels of hierarchies (maybe they don’t want products broken down by supplier brand, for instance), and how to change where totals are shown. Most BI tools have these capabilities, but users may need a simple introduction to use them. Additionally, make this training available to new hires that come into your organization, either through recordings or simple, accessible how-to documents.

2. Explore Excel “cube formulas.” One of Simba Technologies’ large banking-industry customers extensively uses cube formulas to produce active reports that have non-PivotTable layouts. You can ask Excel to “Convert to Formulas” (see OLAP Tools in the Options ribbon tab), which makes each measure cell in your PivotTable into its own OLAP query. You can then insert rows and columns within the resulting cross-tab table, delete cells, or move cells into a very custom layout that is appropriate for your analysis.

This banking-industry customer has become so familiar with the few basic cube formula elements that it now uses them to create approximately half of its new reports from scratch, rather than starting from a PivotTable.

We’ll add two additional notes about this technique and a usability suggestion. First, once a PivotTable is converted to formulas, you can’t further drill down or slice it. However, it does not become spreadsheet-like and stale because it is still refreshable. Second, you would not want to use this technique with a PivotTable that contains thousands of cells because each cell becomes its own query. The technique is more useful for refreshable, custom, high-level reports and dashboards.

Finally, the cube formulas themselves can be cryptic, especially if your dimensions, hierarchies, levels, and members are not well named in the underlying OLAP source. If a user who has been sent this spreadsheet clicks on a cell, it is harder for them to understand what the data in that cell really is. This ambiguity is in contrast to a data cell in a true PivotTable, which has clear columns and rows, as well as column names and row names. One interesting way to get around this problem—and problems of difficult names in the underlying data source—is discussed in detail in a blog posting by one of the authors at http://blogs.simba.com/simba_technologies_ceo_co/2011/05/financial.customer.leverages.novel-dashboard-tables.html.

Tip #4: Improve performance with filters. OLAP PivotTables in Excel and other BI tools give users great power. However, some of Simba Technologies’ customers say that performance issues will be an issue for any system if users don’t realize “how big” a question is that they are asking of the OLAP server. For example, consider the challenge of large dimension hierarchies with, for example, 100,000 members. This size (and even larger hierarchies) can occur when an organization such as a bank decides to put every customer account into a dimension hierarchy.

Hierarchies of this size (and even bigger) usually are not a problem by themselves; this size is what OLAP systems are good at. However, if you place the cross-join of a 100,000-member hierarchy with perhaps a 1,000-member product hierarchy on the same PivotTable axis at the same time, you are asking the OLAP system to consider 100,000 times 1,000 (i.e., 100 million) result rows at the same time. It is common that most of these PivotTable result rows are conceptually empty because not every customer account bought every product. In fact, this sparseness is a characteristic of most OLAP systems. It is worth training users about situations that can lead to multiplicative excesses. For example:

- Teach users to apply/drag the first hierarchy to the PivotTable axis and then immediately filter on that PivotTable axis down to the members of interest to users. In keeping with the above example, do users really want to consider all 100,000 customers? Can they use Excel’s “keep only” feature to eliminate the customers by location, account number range, or any other mechanism? Only then should they consider dragging a second hierarchy over to that axis.

- Excel has a “Defer Layout Update” checkbox at the bottom of the PivotTable Field List pane. Temporarily checking this box allows the user to create the
PivotTable, assign hierarchies to row and column axes and report filters, and then filter some of these dimensions down to reasonable numbers before actually issuing the multidimensional query. Excel is wonderful for dragging hierarchies to axes and filters and instantly showing the results in real time, but sometimes you want to defer the query until you design your layout and reduce “multiplicativeness” via filters. Note that you should always clear this checkbox before closing Excel.

- Make sure that your OLAP data source is patched. For example, adding multiple N of M filters (where \(N\) is 2 or greater) to the report filter (sometimes called page filter or page slicer) in the OLAP option of the Oracle database can improve filter performance.

- As an OLAP administrator, strive to design your dimensions, hierarchies, and levels so they align with your business. For example, avoid a large, flat product dimension. Instead, break down such hierarchies into levels for divisions, brands, departments, categories, or several of these, in ways that make business sense. These breakdowns will allow end users to drill into each branch of a hierarchy a bit at a time and will avoid generating huge queries against your data store. In addition, good hierarchy modeling may allow you to easily apply elegant, hierarchy-branch-based security.

- Avoid hierarchies where a parent node may have a huge number of children. Model or break down your hierarchy into levels that both avoid this huge grouping and that are meaningful to your business users. If you can’t, as when a hierarchy is composed of a very large number of customer account numbers, you can introduce an artificial breakdown. In our bank example above, 100,000 customer accounts can be broken into ranges 00000 to 09999, 10000 to 19999, and so on. In spite of this extra level, a user drilling down to find a particular account can easily navigate to it. The response time will be good, and you will avoid performance complaints to your OLAP administrators.

- It can help to have an “all” member at the top of each hierarchy (such as “all customers” or “all products”), rather than multiple top-level members, and to make the “default” member of each hierarchy that “all” member.

- If possible, push complex calculations (as, say, calculated measures) to your OLAP server, where your horsepower is.

- Teach users that they should not “copy” Excel PivotTables and PivotCharts into Microsoft Word and PowerPoint. Instead, “link” to the spreadsheet element from other Microsoft Office documents so that they too will be updated when the spreadsheet data updates.

**Conclusion**

BI is evolving away from the static monthly or nightly reporting model. Through the use of modern OLAP, multidimensional data sources, and tools for real-time analytics, BI is now more dynamic and ad hoc. The reality is that Excel is a tool of choice, enabling pervasive BI for many users. By appropriately putting the power of self-serve BI in the hands of more users, you can better tune and enhance your organization’s performance. In addition, by following some of the tips we presented, you can eliminate stale data spreadmarts, tie all analysis to one version of the truth, and improve query response times for even your casual ad hoc users.
Agile Project Leaders Redefine IT (Innovative Thinkers)

Valena Hernandez

Abstract

Agile is a project management technique designed to be customized to fit the development project at hand. It answers the challenge posed to IT to deliver more in less time and with proven business value.

Agile projects follow a formal methodology and process. This article presents a classroom example of how agile represents a natural thought pattern that was born out of necessity to meet the demands that IT faces. We reveal ways to uncover that pattern and integrate it into an agile project—a change in thinking that is fundamental to a successful agile project.

This article is not intended as an introductory lesson on the formalities of agile development. Instead, it highlights the behavior patterns that IT struggles with when presented with a discipline that demands creative thinking and acting outside of the box. We do not deny that it takes a strong, agile leader and a commitment from management to embrace the discomfort that comes with breaking down personal barriers, engaging in new ways of thinking, and accepting new methodologies.

We emphasize the IT talent pool as a natural place to look for an agile leader because it helps us identify traits often found in successful agile leaders. These traits can be transformed into good, agile management skills by the necessity to deliver a project on time and within budget. We propose that an experienced agile leader can groom this resource to become a valuable corporate asset. For the purpose of this article, we affirm our success with agile project management techniques to successfully implement both business intelligence and data governance projects.

An Agile Exercise

At a recent TDWI BI Executive Summit focusing on agile development, each speaker described a real-life project. The speakers clearly explained the project's
objectives, the business and technical environments, and corporate culture, and hinted at some of the project challenges. After each speaker’s presentation, the attendees at various tables were given 15 minutes to develop a recommended plan of action that would address the project’s challenges. After considerable table discussion, a representative from each table presented that table’s recommendations. The speaker then described what was actually done and compared it to the attendees’ recommendations.

It was pretty straightforward—or so we thought.

As you can imagine, most attendees were employed in leadership roles, so there was no shortage of leadership. In fact, the first exercise required attendees to strike a delicate balance between choosing a leader who would lead to success versus choosing a leader who assumed he or she was the best person to lead. Eventually, we completed the first part of the exercise when the group effectively assigned the leader by saying, “Hey, why don’t you be the leader since you are taking the notes?” It may not have been the right decision, but by that point we were at least five minutes into the exercise; we didn’t have more time to spare.

Once we selected a leader, we could dive into the problem. Participants leveraged their own experience to determine which of the facts presented by the speaker were important to consider, but our different experiences—our successes and failures and our varying corporate environments, for instance—quickly created challenges, too. We all tried to practice the meeting and group protocols that executives use in heated discussions, but we only had 10 more minutes. Ultimately, the clock forced us to agree on a list of facts pertinent to the exercise.

Next we took a closer look at the facts. We had a list—but we did not agree on what was most important, and each participant gave another burst of decision points and pitfalls. We were at a critical juncture where throwing out a point too quickly could come back and bite us, but spending too much time on it would mean falling short of the deliverable. That conundrum became a discussion point as well: was it better to have a few valid points that made a strong case for success or to have many points with ample padding so it was clear that we considered everything?

Time was short and the protocols were there, but without a leader, we would not be ready. The participants calling for action exchanged subtle communication signals, and finally, the group presenter took the lead and mitigated the discussion points entertained mostly by one participant. Was it the right move? Did it mean that the one participant was wrong? Given the time constraint and pending deliverable, we were all under pressure to move forward without much concern for whether it was right or wrong. No one felt confident in our approach or analysis, but we all knew we gave it our best shot and learned more about each other in 15 minutes than we probably would have in a six-month project. All in all, we spent about eight minutes sifting through the facts and determining which were pertinent and which were given just to see if we would go off track.

The clock was counting down, and the rush of an approaching deadline consumed us all. We tried to think straight for that last step of formalizing the presentation. We hurried to reach a consensus on the order to present and kept adding to the list until we were instructed to stop. Our leader whispered a promise to do his best, but then confessed that, in all the heat, his notes were “sketchy”; he reassured us, though, that he would do his best to deliver the material. We exchanged subtle, nervous signals around the table once more.

Our group was the first to present its analysis. Much to our amazement, the room accepted our analysis as credible and thorough, and we were applauded. What better confirmation did we need? We were on a roll.

It only got better in the next group exercise. More people joined our table—only two people, but it was a morale boost. An original participant from our table—a volunteer with confidence—decided he would like to be the group presenter, which created an air of actual excitement. We were off to a quick start. We spent less time jockeying for a leadership position and even less time
trying to agree on a process to quickly ascertain the facts. However, we still wrestled with the facts and the bearing they had on the solution.

We had the same time constraint as the first exercise, but a sense of calm prevailed because each participant quickly and jointly agreed and/or disagreed on what was important to discuss and we were able to set the priorities for deeper analysis. Because we were able to spend more time on the analysis and the pros and cons, our notes were more direct and organized—and what a difference it made. The deliverables seemed like less of a focus and instead we each openly acknowledged how nice it was just to work together. We were having fun!

**Agile Parallels**
The commonalities between the Executive Summit group exercise and an actual agile project begin with the baseline “project” structures. Both have set time periods that are not open to extensions. Each has a scoped set of deliverables that are expected to be as complete as the time frame permits, so although the speaker expected an analysis of the stated problems, he also acknowledged that the short time frame could affect the level of completeness. Likewise, agile projects are designed to complete deliverables incrementally. Finally, the resources in both types of projects become acutely aware of the pressure to perform in a short period of time.

Each case needs a project leader. Project leadership is often assigned to individuals with management skills and/or who are trained in the discipline. Typically, this person knows how to raise red flags when pertinent, can resolve road blocks, and may have experience with project planning software. This is a short list of project manager attributes and in no way means to trivialize the position, but these skills are the major ones used to meet the deliverables in the assigned time frame.

An agile leader’s skill set is not much different from that of the traditional project leader, except that the former’s success is measured only by the ability to deliver on time. It is a given that red flags are indicators of the need to find alternate solutions that will not impact deliverable deadlines. These issues can then minimally impact the quality of the deliverable.

In our exercise, the group spent the first part of the meeting determining which participant had the wherewithal to guide the group to meet the deliverable in the given time frame. We didn’t have time to discuss our credentials but did our best to intuitively make a selection. An agile leader has the confidence of the project team if they identify that leader as someone who can deliver under pressure and on time.

We had to submit to a leader out of pure necessity, because wrestling with the politics cost too much time and jeopardized our ability to deliver on time. Did it mean everyone was happy with the choice? I don’t think so, but it was fine; it’s simply how we felt after the first five minutes of working together.

Having a leader certainly is a first step in organizing a project team, but there is still no project without the team. We all know that team dynamics are a key factor to success, whether agile or not. As a technology group, we are somewhat challenged even before we start to pull together a team. Technology is a broad subject with many areas of expertise, and corporations commonly organize each area of expertise as a department within the IT organization. Within the realm of our IT disciplines, however, we are quite a diverse group of people, and certain personality types do better in one department than another. Additionally, within each department we also have a range of expertise based on years of experience and exposure to various business and technical environments.

As a project team, we are pulled together from across departments to deliver a project and work as one team. Our classroom exercise released us from all of the project logistics, but we shared the same project goals to review requirements and arrive at an implementation that would be best for the company. In 10 minutes we defined pertinent facts as information artifacts that directly impacted the deliverables, discussed only pertinent facts because time was valuable, and tabled all other topics (to be discussed as time permitted). Our conversations had to
be direct and to the point. All in all, if we didn’t behave as a team, we would not finish the exercise.

Granted, our corporate work environments have their own set of challenges that the summit couldn’t possibly simulate. For instance, we manage departments. We have to balance departmental goals, department budgets, and overallocated department resources. We have meetings upon meetings to arrive at conclusions that everyone is comfortable with. Even when we execute a project solution in our department, we may have to obtain a second department’s approval, whether or not they have an assigned resource to the project. In this instance, when we say that we are all on the same project team, we mean that we share the same goals, but there are still external considerations. No doubt, department structures will always exist, but agile leaders must manage resources to behave like one department and have executive-level support to counter “red tape.” Our classroom environment provided such a scenario.

When we presented our case analysis, the time constraint necessitated an incremental solution. Even so, it was difficult for the participants to agree on the core deliverables. There we were, executive-level colleagues, wondering if our solution was on target. We took shortcuts in our analysis, shortcuts in arriving at conclusions, and shortcuts in preparing to present. It made us very uncomfortable, even though it seemed we were following agile principles as directed. However, it didn’t seem right to speak out and possibly offend someone.

Agile leaders face these same team dynamics. Shortcuts are called shortcuts for a reason; the term suggests that something is missing, and it’s no wonder that the team had some apprehension. Agile leaders must have a can-do attitude while being technically savvy enough to execute solutions that lend themselves to serving as the foundation for the next iteration. The agile leader must be able to envision the end result and build to that end state. Unfortunately, more often than not, they are alone in their convictions.

Team dynamics are, after all, a collection of actions and reactions based on experience. It is not the specific intent of an agile leader to earn a team’s trust and confidence, but it is a by-product of a successful iteration. In our classroom exercise, it wasn’t until we had some level of success that the participants had confidence in the agile leader and each other. The exercise demanded we work out our differences and focus on a solution to the problem at hand, but it also set a thought pattern for us to follow in the remaining exercises. Ideas were openly entertained and discussed when viable, or put to rest without consequence if deemed out of scope. It became a comfortable process born out of the necessity to deliver.

If we adopt this thought pattern, we are all candidates for becoming agile leaders.

A Final Word
What are your agile qualifications? Have you ever been in a project kickoff meeting after which people were quietly discussing all the reasons why a project will hit roadblocks, but are offering no solutions? Are you the person who knows the roadblocks but can’t seem to get anyone to act on resolving the issues? Do you wish you had a leader who could sift through department resources to just get it done, or are you the point person who is always asked to go under the radar to complete a task that, on the management level, could not be done?

Being agile is all about execution. An agile leader has the satisfaction of knowing that he or she is making a tangible difference to the company and helping to transform the thought leadership of future projects.
BI Experts’ Perspective

Moving to Real-Time BI

Sami Akbay, Harriet Fryman, David Stodder, and James Taylor

Cookie Delight manufactures, sells, and distributes upscale cookies throughout the U.S. Its business intelligence (BI) director, Randy Nix, has done a great job moving the company along the BI maturity curve. There is an enterprisewide data warehouse, a comprehensive reporting system, dashboards and scorecards, and analytics to understand and predict customer behavior. There is now an analytical culture at Cookie Delight.

The company has manufacturing facilities in Tampa, Austin, Portland, Chicago, and Baltimore. Sales representatives call upon both grocery store chains and “mom and pop” retailers. It has an enterprise resource planning (ERP) system for financials, human resources, and manufacturing, along with a number of specialty legacy applications. The latest upgrade to its data warehouse lets employees capture and store real-time data.

Randy knows that many firms are now doing real-time BI and that some are automating operational decision making, but he needs to know more before introducing senior management to this new direction. In particular, Randy wants to know:

1. What are some of the “sweet-spot” applications that other firms like Randy’s have found for real-time BI, and especially for automating decision making?

2. Is moving to real-time BI disruptive? Is it just a further BI evolution, or is it a fundamental shift from what is currently being done at Cookie Delight?

3. How would the movement to real-time BI impact his BI staff? Are staff members likely to have, or easily acquire, the necessary technical skills?

4. What new technologies are required? How do rules engines fit into all of this and are they required?

5. How do you get employees engaged, excited, and ready to use real-time BI information?

6. What are some of the “gotchas” that other firms have experienced in moving to real-time BI?

7. How do you recommend that Randy move forward?

Sami Akbay is CEO for Altibase, Inc.
sami@akbay.com

Harriet Fryman leads product marketing and market strategy for the business intelligence portfolio at IBM.
hfryman@us.ibm.com

David Stodder is TDWI Research Director for Business Intelligence.
dstodder@tdwi.org

James Taylor is CEO for Decision Management Solutions.
james@decisionmanagementsolutions.com
Real time is the key enabler to operational analytics. Traditional BI empowers strategic decisions and helps business analysts and management. Real-time BI augments the traditional framework and allows businesses to improve tactical decisions for operational functions. The shift to real time helps operational staff figure out “what they need to do now,” supported by the intelligence of their strategic analytics infrastructure.

Randy is in an admirable situation—Cookie Delight clearly has an appetite for analytics since the company has already deployed reports, dashboards, and scorecards. Randy has an opportunity to bring more of what people already like and to empower more people at Cookie Delight with BI by making it operational.

Using real time to make BI operational will change the profile of Randy’s end users. In addition to management and business analysts, Randy will need to cater to the front-line employees who are in the field. The road to success and acceptance starts with a clear understanding of this new profile and winning over the operational staff—in addition to his traditional internal customers.

Real-time BI requires changes to the IT infrastructure and the BI software portfolio. The BI team will need to acquire new skills, and business processes will require revisions. Gradually evolving toward real time would be less disruptive than making big and abrupt changes. Randy would be taking too much political risk if he decides to be a revolutionary. As with any change, success will require some amount of trial and error; Randy needs to carefully pick and choose where he can take technology and process risks.

Delivering real-time BI to a new group of employees will have the biggest impact.

Randy should divide the project and conquer small battlefields to win the war. He should start with less complex and more impactful “mini-projects” first. This approach serves multiple purposes: instant gratification for his internal champions, confined exposure of his team, and a reduced-risk test bed for infrastructure changes. By the time the first project is rolled out, the team will have some hands-on experience, unexpected infrastructure requirements will surface and be addressed, and new packaged BI software will be tested in a real-world environment. Of course, shorter time to delivery will help maintain internal support.

Delivering real-time BI to a new group of employees will have the biggest impact. Although the analytical culture at Cookie Delight means that the employees are more likely to embrace real-time BI faster, change always meets resistance.

Internal marketing is essential and it will positively impact the adoption of real-time BI. Training end users for the new technology and convincing them of the benefits should never be overlooked. Coupling training with incentives goes a long way, especially when you’re changing some deeply rooted business practice. Your traditional internal champion in management ranks may get a bigger bonus when his department contributes more to the bottom line; however, your operational end user may be too far removed from perceiving this bonus as an incentive. You need to convince your end user that he or she is better off using this new system.

End-user buy-in is important and Randy can try to get creative; for example, if Randy can get “upgrade cycles” for end-user equipment to align with his BI rollout schedule, he can benefit from the enthusiasm that goes along with a new tablet, smartphone, or PC that is distributed to the end users when the real-time BI application is deployed.

Randy’s team will certainly be impacted with the move toward real-time BI. The new portfolio of BI tools and its associated learning curve can be a challenge, and Randy may need to bring new staff members or consultants into the team.
Supporting real-time BI applications requires tight cooperation with IT operations. For the infrastructure, getting continuous data from operational systems is different from receiving batches of data at convenient intervals. Operational analytics comes with an expectation of “always on” and “always available.” Performance becomes paramount; a business analyst who spends 30 minutes waiting for a report will not be happy. When an operational system takes two minutes to respond, people will stop using it.

Automation should be introduced gradually. Rules engines and event-driven analytics may often improve efficiency, but unless the impacted business processes are reviewed and revised to accommodate automation, things will inevitably go wrong. Automation should first be introduced where end users feel that a task is repetitive, menial, and tedious. Especially in the beginning, the end user should not think that he has given his decision-making authority to a rules engine. The new BI framework should help the end user improve how he does his job, not remove the end user from the job.

In summary, Randy needs to understand his new end users, deliver smaller projects at first, provide proper incentives and training to ensure internal acceptance, and automate slowly. He then needs to repeat this process as he continues to deliver more real-time BI. Getting to real time is an ongoing process—not a single monolithic project.

Harriet Fryman

It confuses me why Randy has already captured real-time data in his data warehouse and preempted his technology choice before he seems to have determined its business application. He seems to have fallen into the common trap that causes many BI projects to fail; they start bottom-up with, “I have data; now what can the business do with it?” With that said, I congratulate him for now taking stock of how he can best serve Cookie Delight with real-time BI, and how the technology and management of his BI project need to evolve for the organization to capitalize upon its value. He seems poised to start this initiative top-down with, “I have found business value for real-time decision making; how can I best serve it?”

Real-time BI augments the traditional framework and allows businesses to improve tactical decisions for operational functions.

To pursue this business-in thinking (versus information-out thinking), Randy needs to take stock of the five core tenets that determine the success of every real-time BI, or operational decision-making, initiative:

1. Where in the organization do decisions need to be made in real time, and which of those processes would benefit the most from more timely decision making?

2. Who makes those decisions, and can any of those decisions be automated with business rules to reduce or eliminate the need for human intervention at every decision point?

3. How can real-time information be best sourced, organized, and made readily available? Is it pulled or pushed from a resident store? Is it in-flight streaming or messaging data? What transformation, aggregation, or comparison to context data is required to make the real-time data meaningful? Does it need to be stored long term or simply served and discarded after the decision is made?

4. Where is the best place to serve up the real-time information? Are the decision makers expecting the real-time data in an on-request report, or are they continuously updating their dashboard where they see their other BI information? Do they expect it distributed to them instantly via e-mail, SMS, or in headlines in their portal? Are they best served with the real-time BI within the business process itself, in a business process management (BPM)
process, or in a transactional application context?

5. When the real-time decision is made, how is it then made actionable? Once decision makers have real-time information, what are they going to do with it (and when)? Will the decision initiate a business process, assign an individual with a task, or instantly replan or reallocate resources? If the decision isn’t in real time, then the real question is, “What is the need for real-time BI?”

By working with the business to answer these questions, I expect Randy to find high-value business processes in manufacturing and distribution that can be optimized with real-time insight. I advise Randy to look at:

- Streamlining supply-chain logistics such as optimizing truck routes; eliminating shipment bottlenecks; minimizing inventory through proactive warehouse management; or monitoring radio frequency identification (RFID) and real-time location system (RTLS) sensor output to reduce spoilage, minimize delays, and avoid shipment error.

- Optimizing manufacturing and production decisions such as eliminating stock outages of raw material inventory against production schedules; addressing any decline in production-line yields; switching production lines to respond to urgent customer requests, down lines, or increased demand due to trade promotion; or predicting equipment maintenance to maintain uptime.

- Excelling at customer service with proactive management such as providing current order status, adjusting schedules to ensure on-time delivery, proactively offering product alternatives to prevent out-of-stock situations, or understanding the buying behaviors from scanner data.

Having found the sweet spot for real-time decisions, Randy’s investigation must turn to the technology and team readiness. The key is to not force-fit real time into the current technology purely because it is the known world. Two areas that require particular investigation are:

- Real-time information that can update resident data or flowing in-flight streaming data. The former requires changed data capture (CDC), likely already familiar to Randy’s team. The latter, however, requires real-time BI technology (such as IBM Cognos Real-time Monitoring) that can listen to streaming data, process it in real time, attach historical context, and provide constantly updated dashboards in real time.

- Business rules engines that can be applied to a mix of in-flight or resident data. These rules engines are ideal for real-time decision making as they filter out expected norms or less urgent decisions, so that only critical information is served to decision makers. For this, Randy’s team must collaborate with the business to determine thresholds, what actions to “code in,” and how the business is to be informed by coordinating with alerts or BPM engines. Understanding rules engines and the business process is critical to success.

In conclusion, I ask Randy’s team to be wary of any phrase from the business that starts, “It would be nice to see x, y, z real-time data...,” because the business value is not quantified. Indeed, Cookie Delight should only move ahead when the business can articulate, “With real-time information, I could do x, y, and z to reduce costs/increase efficiency/raise customer satisfaction.” It is only when the outcome of a real-time decision is clear that Randy can be assured his team is providing the organization measurable business.

David Stodder

“Real time” is hot in BI, analytics, and data warehousing circles, in part because the promise of instant answers and instant gratification is always a strong lure. Since its origins, the computing industry has been hooked on speed; faster is always better, and so “real time” must be the ultimate. Real time is exciting, but it is complicated and can be expensive to deliver and sustain. “Real time” also means different things depending on the users and the circumstances, so expectations need to be managed carefully.
A major reason why “real time” has long signified The Promised Land for BI, analytics, and data warehousing is that most of these systems have been stuck in the past, delivering rearview snapshots of business and financial performance, customer trends, and other topics of interest. Historical analysis will always be important, but organizations such as Cookie Delight can gain considerable business advantages if they can monitor and measure the present. One clear advantage is that they can begin to see, in real time, whether their efforts at shaping the present (through actions guided by predictive analysis and forecasting) are having the desired effect, and if not, they can adjust models and strategies sooner.

The major sweet spot for real-time BI and analytics is where conditions are highly dynamic.

Real-Time Sweet Spots

In the Internet age, competitive advantages are ephemeral. New products and services—not to mention cookie recipes—are soon copied or go out of style, which means that organizations must have a timely view of their markets and be prepared to adjust strategies quickly. A more timely view will improve response to customer behavior and help the organization avoid wasting supplies and resources on the wrong products.

Thus, the major sweet spot for real-time BI and analytics is where conditions are highly dynamic. Web site storefronts offer the best examples of such conditions; organizations are under pressure to shape customer experiences by putting appropriate content, marketing offers, and services in front of them in real time. Organizations also need to keep an eye on online social media to monitor sentiment and brand reputation. Certainly, the more organizations are able to view social media data in real time, the faster they will be able to craft the right responses.

Analytics are extremely important for interpreting and predicting customer behavior and then employing various Web tools, applications, and services to execute actions—all of which are designed to encourage customers to spend money, stay loyal, and share what they see on sites. Randy mentions that firms are automating operational decision making; such decision management systems link predictive modeling and analysis with rules-based systems for execution, enabling organizations to increase the speed and efficiency of response. Online marketing and Web sites are good places to implement decision management, among others.

Along with speed, “agility” is an important objective in dynamic business circumstances. Organizations need flexibility to shift strategies, redeploy labor resources, and take actions when conditions change. Real-time data and analytics are important to agility, because obviously the sooner an organization knows what is happening and can make changes, the better. However, organizations have to be careful not to overuse real-time data alerts. Too many alerts become noise and can confuse decision makers, or cause them to ignore information, like the boy who cried wolf.

Good Disruptive or Bad Disruptive?

“Operational” BI, or the expansion of BI systems out to managers and employees in customer service and other lines of business and front-line operations, has already put the pedal to the metal on data timeliness. These employees need to be kept up to date on the status of product sales, supply availability, and more; operational BI deployment has cut down the data latency in many organizations to updates several times a day, if not more than once an hour. Operational data stores, data federation middleware, and other data warehouse and integration technology play a critical role. By upgrading its data warehouse to let employees capture and store real-time data, Cookie Delight will crank its update frequency even higher, which will put pressure on technology systems to keep pace.

When business processes are running with more timely data, good disruptive changes can occur. Organizations have information that allows them to run more efficiently,
serve customers more effectively, and be smarter in supplier relationships, procurement, and other areas. Real-time data will increase awareness, enabling organizations to move even faster to capitalize on opportunities to reduce costs or make process changes.

However, introducing real-time data will demand that organizations ensure that operations that don’t use this data stay in sync. In addition, organizations must make sure that metadata and higher domain definitions are not splintered by the inflow of new, real-time data. Without attention to these information management issues, real-time data could turn into a negative disruption.

Success Factors for Real Time
Real-time data and analytics put new and different demands on BI and data warehouse systems as well as personnel. Rather than assuming that current personnel will be able to handle the requirements, organizations must assess whether they are ready. Here are three considerations that are important for organizations to move forward with real-time data successfully:

- Not all the data needs to be real time all the time. BI directors and data warehouse managers need to guide users to focus requests for real-time data on metrics or updates that demand this level of timeliness. Exception-based reporting and performance metrics could be good vehicles for providing real-time data. Look for dynamic business situations as the best areas to apply real-time analytics.

- Evaluate technology platforms and expertise for availability and reliability. Organizations need to determine what “real time” will really mean. Will it change how dashboards are refreshed or how frequently data is analyzed for patterns or anomalies? The answers to these questions will impact underlying technology platform requirements and whether personnel have the experience they need.

- Don’t assume that application business rules can handle real time. It is often difficult enough to get data out of ERP and other big applications; if update frequency speeds up to real time, business rules written into applications and business processes may need to be rewritten. Make sure developers are able to synchronize rules in applications and processes with real-time data requirements for BI and analytics. This can turn into a critical “gotcha” that can frustrate business ambitions for real time.

Cookie Delight may not yet have to deal with the tsunami of real-time data coming from sensors, tags, and mobile devices. However, organizations trying to leverage this “big data” have had to look outside traditional data warehousing to Hadoop and event processing systems. Real time is definitely a catalyst with the potential to fundamentally change BI and data warehouse architectures. However, Cookie Delight would be wise not to bite off more than it can chew, if you’ll pardon the expression. The company will have time to evolve toward bigger dreams for real time.

Make sure developers are able to synchronize rules in applications and processes with real-time data requirements for BI and analytics.

James Taylor
Finding the Sweet Spot for Real-Time BI
The best way to find sweet-spot applications for real-time BI is for Randy to identify the decisions being made in his operations and to map those decisions to his business objectives and key performance indicators (KPIs). High-volume, repeatable decisions—the kind made every day—that impact important KPIs are the right place to start. Improving these indicators will move the dials that matter to Cookie Delight and will take advantage of the multiplicative effect of lots of decisions, where even a small
improvement gets multiplied for a big total impact.

If getting deliveries to retailers is key to the business model, then focusing on decisions that affect on-time delivery might be best—for instance, decisions about what shipper, truck, or routing to use; how to schedule production to manage inventory; and more. Perhaps cost control is number one, and decisions about supplier selection and scheduling to avoid waste and replenishment will be central. Only an understanding of the business can help you prioritize.

The Disruption of Real-Time BI
The most disruptive element of real-time BI is the move to automated decision making. Presenting data to people gets incrementally more difficult as you move closer to real time, but it is not truly disruptive. However, to make maximum use of real-time data, you will need systems that can make decisions on that data in real time. The move from decision support systems (those that help people make decisions) to decision management systems (which make decisions on your behalf) is a major shift. Policies and regulations that used to be applied by users must now be embedded in the system in a way that is easy to change so the system remains agile.

Analytic interpretation that used to be done by people based on visualizations and query tools must now be done using executable predictive analytic models so the system becomes analytic. Learning no longer takes place outside the system, with people getting smarter about what works. Now it must take place in the system so that the system itself becomes adaptive.

Impacts to BI Staff
Moving to real-time BI will require two big shifts within the BI team. First, they will need more skills for building predictive analytic models and performing data mining. These techniques produce outcomes that are more easily automated and more easily embedded in systems. Visualization, dashboards, and query tools are all great for people trying to gain analytic insight from their data, but they don’t work for systems. Systems need rules and formulas derived from the data and that can be executed.

The most disruptive element of real-time BI is the move to automated decision making. The second change will be a need for more systems know-how. For instance, if a center of excellence is moving from a focus on BI and analytics to decision management systems and analytics, it will need greater IT skills to make sure that the results of analytic work can be fed back into operational systems and drive decision making. Existing systems skills that are focused on getting the data out of operational systems for analysis must be complemented by those for pushing analytic decision making back into operational systems.

New Technologies and Rules Engines
Two new technologies will be critical for moving to real-time BI. First, there is a need for new tools to perform data mining and predictive analytic modeling. It sounds as though Randy’s team is already doing some predictive analytics, so these new tools may not be too different. They will need to focus on workbenches or tools that support what I call an industrial mindset when it comes to predictive analytics—tools that ensure a group of modelers can produce, validate, deploy, and update a portfolio of models in a repeatable, reliable, industrial way.

The other key technology is a business rules management system (BRMS) that gives Randy two essential capabilities. First, it gives him a way to effectively deploy his predictive analytics. Many predictive analytic models can be readily deployed as business rules, becoming more manageable in the process. Some BRMSs even support automatic conversion using a standard known as PMML (Predictive Model Markup Language) to define the models. Second, a BRMS lets his team define what actions should be taken based on the data and the analytics. Real time means making decisions, not just presenting
analysis, and business rules are great for defining which actions to take (and when to take them).

Engaging Employees with Real-Time BI
With real-time BI, you need to get employees engaged with systems that make decisions, not present information, and that’s a different thing altogether. Working on what I call the “three-legged stool”—business, IT, and analytics—is essential here. Only when all three groups are working well together will BI systems that make decisions in real time be effective.

A BRMS can also help. It makes the system’s decision process transparent—business users can see how decisions are being made and add their own expertise, policy makers can see that their policies are being followed, and everyone can see the rules that were derived from the analytics. People don’t like black boxes; transparency really matters.

The “Gotchas” of Real-Time BI
There are several, but three key “gotchas” come to mind:

- Predictions are usable probabilities, not absolutes. A predictive analytic model cannot tell you which deliveries will be late—only how likely it is that a delivery will be late. Such information is tremendously valuable but not infallible.

- These systems often change which users are good or bad at a job. Real-time decision-making systems handle much of the day-to-day transaction volume, so someone who was efficient may have less to offer and someone who was effective may shine because such effectiveness will still be valuable.

Only when business, IT, and analytics are working well together will BI systems that make decisions in real time be effective.

Moving Forward
Begin with the decision in mind. Find the decision(s) that matter, the ones that will move the dial for Cookie Delight. In particular, focus on day-to-day, high-volume, operational decisions. Concentrate on improving those decisions, on making the decisions in real time, and on using analytics to make those decisions as accurately as possible.

Also, bring business rules and a BRMS into the analytic technology mix because it will help make deployment of analytics, and automation of decisions, practical.
Harnessing the Power of Social Media Data with Master Data Management

Ravi Shankar

Who is your most valuable customer? Is it the one who buys the most from you or the customer who has the most influence on other buyers via social media—in effect a customer who doubles as a salesperson?

The value of customers who endorse, or even evangelize, products and services on social media is on the rise. These consumers are a powerful viral marketing force that exerts influence across an extended network of their friends, and ultimately friends of friends of friends. With an estimated 50 million tweets and 60 million Facebook updates a day, social media influencers have emerged as a uniquely valuable customer slice that smart marketers are looking to cultivate.

Cultivating those customers requires visibility into the dynamics of social media data. Companies increasingly recognize social media’s potential to creatively disrupt marketing as we know it and are strategizing over the technologies and processes they need to capitalize on the opportunity and hear the voice of the customer amid billions of messages.

Data is the key to unleashing the value of these customers. Organizations need to capture, integrate, and analyze social media data, a type of customer-interaction data that makes up a sizable part of the “big data” explosion confronting enterprises. A key focus area is the one-to-one marketing that is made possible by understanding spheres of influence and targeting active social media users.

A Lesson in Word-of-Mouth Social Marketing
Consider an example of the power of word-of-mouth marketing from a U.S.-based department store. This large
A retailer offered women free in-store makeup application sessions. The idea was that women who sat for a consultation would purchase cosmetics during that store visit, but the results were disappointing and the program was scrapped. Then came the twist—overall cosmetics sales fell noticeably. Why? After research, the store determined it had lost the word-of-mouth value of its program.

By harvesting social media information from Twitter and Facebook, the store gained a new understanding of the influence model for cosmetics. Customers receiving free makeup sessions weren’t necessarily buying cosmetics, but their word-of-mouth influence prompted purchases by friends and friends of friends. The retailer enriched its customer data with social media data, more precisely targeting influencers with the right offers and increasing its profits.

Similar opportunities exist in many industries—travel and hospitality, financial services, telecommunications, business-to-business (B2B) sales verticals, and others. In all industries, harnessing social media data can help companies answer questions such as:

- Which of our customers are active on social media?
- What are customers saying about our brand and products?
- With whom do they have relationships, and how strong are they?
- What is the extended sphere of influence among our social media customers?
- What common characteristics are shared among social media users?

**Job #1: Achieving a Complete Customer View**

Social media data can be used to enrich existing customer profiles. Suppose a travel and hospitality firm could learn from a Facebook fan’s posts that her family is planning a Caribbean cruise in six months. Suppose a bank could learn that a teenager is looking for a college and exploring student loans. Imagine a B2B technology provider learning from a LinkedIn forum that a customer is dissatisfied with her service and is considering other vendors. Clearly, social media opens a new frontier of marketing possibilities.

Social media data offers unique insights into customer likes, dislikes, plans, history, and networks that can be gleaned nowhere else. The problem is that unless you already maintain a complete and reliable customer profile, the value of social media data is severely constrained. At worst, attempting to harvest and utilize this intelligence generates yet another data silo that complicates your view of the customer.

Despite investments in customer relationship management (CRM), sales force and marketing automation, and other customer-oriented technologies over several years, many organizations still lack a complete view of their customers, ranging from reliable and consistent names and contact information to the products and services in place. In both business-to-consumer (B2C) and B2B, customer data is often fragmented across CRM, enterprise resource planning (ERP), service, and other applications. Sales, marketing, and service personnel waste time piecing together customer information from various applications.

For instance, in the U.S. retail sector, 92 percent of respondents to an Aberdeen Group survey reported that they do not integrate customer data from all their marketing channels. The same study found that 77 percent of companies characterized by Aberdeen as “laggard” don’t have a multichannel database marketing solution in place (Cunnane, 2010).

**Master Data Management: A Key to Customer Centricity**

In recent years, more organizations have been turning to master data management (MDM) to solve the problem of integrating customer data. Using MDM to generate a single, authoritative view of the customer by reconciling and consolidating disparate data from multiple systems is the top application area for MDM—ahead of supply chain and enterprise operations.
MDM has been called the “secret sauce” for CRM 2.0, a centerpiece of which is social media data. Supported by data quality and data integration, MDM is used today by organizations around the world to realize the goal of customer centricity—improving customer acquisition and retention, increasing revenue and profitability, and enhancing business speed and agility.

Today’s best MDM solutions allow you to generate progressively richer customer views as your needs dictate (Figure 1):

- **A single customer view.** Names, addresses, e-mail addresses, phone numbers, and other identifying information is consolidated and reconciled into a single customer view. The solution applies data quality deduplication to address discrepancies, can offer users choice in which attributes to use, and ultimately produces a “golden record”—the best version of the truth.

- **A 360-degree view of customer relationships.** The single view is expanded to incorporate the customer’s products across lines of business, positioning the company to execute informed cross-sell and up-sell campaigns. This view also adds the customer’s household and business relationships; it then lends itself to engaging a customer’s family members and thereby building multigenerational value, or selling into the customer’s business.

- **A complete view of customer interactions.** The complete view expands on the previous two levels by adding data on customer interactions and transactions, such as support activity, payment status, billings or subscriptions, or warranties and registrations. It can reflect customer activity on social media (e.g., data from Facebook, Twitter, YouTube, and other sources; reviews or comments written on blogs) and activity in channels such as online storefronts and call centers.

Figure 1. A progressively richer view of customer data.
Under the MDM Hood

From an architectural perspective, MDM functions in a hub model that captures disparate data from select sources into a central MDM repository. Prebuilt and customizable algorithms are applied to reconcile, deduplicate, and consolidate this heterogeneous data into a “golden record.” Continuing maturation of MDM technology has given rise to the concept of “universal MDM,” which describes managing all master data regardless of its type or location with a single, unified solution. Four key characteristics define universal MDM:

- **Multi-domain:** Master data on customers, suppliers, products, assets, and locations can be managed, consolidated, and accessed.

- **Multi-style:** A flexible solution may be used in any style—registry, analytical, transactional, or coexistence.

- **Multi-deployment:** The solution may be used as a single-instance hub or in federated, cloud, or service architectures.

- **Multi-use:** The MDM solution interoperates seamlessly with data integration and data quality technologies as part of a single platform.

“MDM-Aware” Business Applications

A recent development of particular interest to leveraging social media data for marketing is the emergence of “MDM-aware” business applications. This capability enables MDM developers to embed MDM functionality directly within the applications that business users access every day, such as Oracle, Salesforce, Siebel, and SAP. The objective is to put complete, trusted master data about customers into the hands of business users.

Working through the business application of choice (or directly with the MDM user interface), the back-end MDM solution will guide a business (a sales representative, a marketing manager, or a service agent) through the steps that lead to a complete customer view. The solution:

- **Automatically detects and reconciles duplicates.** The MDM solution automatically detects duplicate records, such as the same customer named John Jones in one application and Jon Jones in another, and intelligently merges them into a single record.

- **Ensures the accuracy of customer data.** Before the data is presented in the CRM application, it is cleansed of contradictions and inconsistencies that likely exist across disparate applications.

- **Presents customer transactions and interactions.** The solution presents customer transactions and interactions (e.g., social media, call center inquiries) that reflect a complete history of customer activity.

- **Syncs on-premises and software-as-a-service (SaaS) applications.** As needed, the MDM solution synchronizes data among any combination of traditional and SaaS applications.

- **Incorporates third-party data.** If appropriate, the solution interoperates with third-party providers to incorporate demographic information into customer profiles.

- **Delivers real-time alerts to business users.** Complex event-processing technology alerts business users to important changes, such as a negative review on a social media forum or a balance falling below a threshold.

- **Displays extended relationships and activity.** The solution aggregates and visually presents data on a customer’s extended account, business, family relationships, and social networking relationships.

Using Social Media Data for Service and Brand Management

Besides targeted marketing, harnessing social media data can enable an organization to respond decisively to customer service and other issues and align its service to customer value and social influence. For instance, tracking a customer’s social media activity can provide an alert when the customer posts a product or service
complaint on Facebook or Twitter. With the right data infrastructure and analytics, you can:

- Assess the customer’s purchase history, past complaints, and overall value
- Determine his friends-of-friends sphere of influence
- Respond proactively as appropriate to the customer and complaint
- Deliver a positive customer experience and mitigate brand damage

Social media is also a new and exciting channel for brand-level marketing and getting your message in front of relevant people. For instance, at Informatica we use Twitter to influence the influencers and engage directly with customers. We write blogs to tell our point of view (and increase search engine optimization [SEO] performance), and we engage in online communities to show our thought leadership and influence customer and partner perception. A few examples:

- An announcement we made earlier this year with NetSuite, the cloud-based business management provider, was retweeted 366 times by different people. Within the Twitter world, people follow other people because they are interested in what they have to say. As a result of these tweets, 253,485 people received the news in their “social media” inbox. This is specific, targeted marketing.

- When we launched Informatica Data Services, we looked at how to deliver our message to enterprise architects. We decided to build an online community within LinkedIn, and over six months we grew this community to over 1,300 professionals interested in this subject.

- We recently augmented our Global Customer Support service through a new Twitter handle as a supplement to our standard support offerings. We are already answering specific questions and providing a useful information service to our customers.

For brand management, social media data provides a forward-looking, proactive channel that reflects interactions between people (compared to the “rearview mirror” we’re accustomed to with traditional, relational data). Brand management requires large-scale statistical sentiment analysis across the broad social media world; marketers can mine that data and understand individual comments.

The way we access all of this data may change given the uproar over privacy and the new bill called “Do Not Track Me Online” moving through the U.S. Congress. However, there is still a significant amount of public information available (blog posts, LinkedIn, tweets on Twitter, etc.). Companies in virtually every industry have the opportunity to harness social media data for brand management.

Informatica estimates that roughly 20 percent of social media data is meaningful to marketing and customer management; 80 percent is irrelevant to the user organization’s interests.

For instance, suppose a hotel chain wants to understand the sentiment of its brand and compare it to that of its competitors. Here there are no privacy concerns because most commentary exists on public sites (such as TripAdvisor). By harnessing this data, the chain can understand how it is judged and drill into specific comments made by individuals about its offerings or its competitors.
Unique Challenges of Social Media Data

Using social media data poses some unique challenges. One is extracting relevance from many millions of messages; another is ensuring that unstructured social media data can integrate with traditional sources. Organizations should understand social media data in a broader, “big data” context and assess capabilities for MDM and data-integration solutions to interoperate with high-performance, distributed-processing platforms such as Hadoop.

Informatica estimates that roughly 20 percent of social media data is meaningful to marketing and customer management; 80 percent is irrelevant to the user organization’s interests. It would be counterproductive to extract and retain every byte of a customer’s social media activity, which would result in data volumes four times larger than necessary.

Establishing business rules to govern key aspects of the process is required to minimize volumes, enhance performance, and heighten relevance. Ideally, the data-integration solution applied to social media content will feature native connectors to such leading sources as Facebook, Twitter, and LinkedIn. Your business rules checklist should include these tasks:

- Identify which customers to follow
- Identify which social media vehicles they use
- Implement keywords to monitor on social media vehicles
- Automate data extraction based on customer and keyword business rules
- Continually monitor and optimize the process

The Social Media Sweet Spot: Data Integration

For purposes of one-to-one marketing, the sweet spot is at the intersection of customer-interaction data (such as social media) and conventional transaction data (an existing CRM profile covering contact information, purchase history, demographics, etc.). Capabilities to transform unstructured social media data, improve its data quality (if necessary), and integrate it with relational data are essential to making the most of the one-to-one opportunity.

Consider this example. If you can create a complete profile that includes a customer’s birthday (from your existing profile) with her status as a Facebook fan, you’re positioned to send her a birthday greeting via Facebook. Your solution should also be able to create segments of your social media fans and followers, laying the foundation to analyze those customers for common characteristics and enable macro-level targeted marketing.

Social Media within the Broader “Big Data” Context

To appreciate the potential of social media data, view it in the broader context of “big data” and understand some of the emerging requirements that big data demands of enterprises that want to harness it.

Social media data is a type of big data that has generated considerable interest, but others offer value as well (varying by industry). Although market definitions differ, Informatica defines big data as meaning all data. Big data comprises internal information in enterprise applications, data warehouses, cloud solutions, storage systems, and mobile or specialized devices. It also includes external sources such as social media, partner applications, and public Web sites offering interactive and downloadable content. Big data can be categorized as transaction or interaction data:

- **Big transaction data:** Traditional relational data continues to grow in transactional and analytic systems, from ERP applications to data warehouse appliances, along with unstructured information. The landscape is complicated as enterprises move more data and business processes to public and private clouds.

- **Big interaction data:** Besides social media data, interaction data includes call detail records (CDRs), device and sensor information, Global Positioning Service (GPS) and geolocation-mapping data, large image files (through managed file transfer), Web text and clickstream data, scientific information, and e-mail messages.
For the last 40 years, enterprises have been comfortable with the relational database model and building applications that leverage this transactional data. Now, because of the ultra large-scale processing required by social media and other sources of big data such as Web clickstream and mobile devices, enterprises are turning to the open source Hadoop platform.

Hadoop offers an attractive platform for storing and managing social media information. Making the most of it, however, requires that an enterprise can seamlessly load data from any source into Hadoop; extract data from it for delivery to any target; and apply data quality, deduplication, and other processes to Hadoop-based information.

We are just beginning to understand the implications of this new world and discover new use cases for Hadoop.

We are just beginning to understand the implications of this new world and discover new use cases for Hadoop and big data within our customer base. Examples of big data in action include:

- **Fraud detection**: A credit card provider needs to drive risk-detection analytics. Their traditional data warehouse is not able to process the data effectively, so they move data from their data warehouse into Hadoop to generate the “machine learning data set” to support risk detection. They then analyze the patterns of fraudulent online behavior and build machine learning solutions.

- **Customer analytics**: A telecommunications company wants to analyze call patterns for customers across its network. Upon investigation, they determine that members of a “friends and family” network have a 25 percent chance of deserting the network should any member leave. The analysis required to understand this trend requires vast computing power enabled through Hadoop.

- **Large-scale financial analysis**: A financial services company wants to analyze the last 10 years of stock market ticker data. They implement a Hadoop cluster so they can perform the vast processing required to analyze this data. As a type of big data that is generating significant interest, social media data offers an epic opportunity for marketers to improve cross-sell revenue, one-to-one targeting, service effectiveness, and brand management. Effectively using social media data demands a powerful, flexible data-integration platform and MDM technology that can incorporate social media data into a complete customer view and help you cultivate social media consumers as some of your most valuable salespeople. Those same technologies may be repurposed for other types of big data and application areas, positioning you to turn big data into big opportunities.

**Reference**
Four Strategies for Increasing Business Intelligence Usage

Khwaja Saad Waheed

Abstract
To get the most from their business intelligence (BI) investments, enterprises need strategies to increase the use of BI. The four strategies presented here are based on more than 10 years of field experience developing and implementing BI solutions for large enterprises. We explain the best implementation approach, examine the importance of drill-down functionality, explore a new concept called action-integrated BI, and discuss how to incorporate metadata to assist users. In addition, we describe why a properly designed security model is important. Finally, we provide specific guidance for a typical BI installation and show you how to avoid pitfalls.

Introduction
Much has been written about how to get the most from your investment in BI and data warehousing programs. One key to success is getting as many users as possible to work with systems that deliver the right amount of detail to the right people at the right time. The problem is that users resist change. A new BI report can create fear and trepidation, which will hinder your goal of making BI pervasive.

This article will examine four strategies for overcoming common problems. In addition, it will examine a tangential issue—data security—that can trip up many implementations.

Strategy #1: Retire Your Legacy Reporting Framework
Retiring old, legacy reporting is essential to the success of a data warehouse (DW) or BI project. Change is always challenging, especially if the change occurs for something that is not part of a person’s core duties. I have seen several truly good projects fail to gain broad usage for
daily operations and strategic planning because users did not adapt to the BI framework.

The most common strategy used to implement a BI project is popularly referred to as “parallel run.” The approach is often justified by claiming that it lets workers use the new technology with the comfort and peace of mind of having the old reports still accessible. Another reason often cited is that it allows users to match the new and old reports so they will begin to trust the reports from the new BI system. The idea is that report consumers can gradually and naturally build confidence in the BI solution so old reports can be systematically retired over time.

In my experience, parallel runs are unsuccessful. A core assumption—that the outputs of DWs, data marts, or other online analytical processing (OLAP)—type solutions must match the legacy reports—is faulty. When two reporting applications are developed with different technology or intended uses, the results will not necessarily match. More often, the underlying business rules are changed in the name of accuracy without anticipating the impact it will have on business metrics and key performance indicators (KPIs).

Furthermore, why would a company invest significant time and resources to build a DW that simply reports what they already know? The selling point of BI and DW projects is to find the unknown, to dig deep and uncover a perspective on the organization that was never before accessible or understood.

Allowing users access to older reports creates a crutch that makes it easy for them to slip back into old habits. It removes any incentive to learn something new. People will endure a learning curve more effectively in a “sink-or-swim” environment. There are always high-priority tasks that seem to call for use of the old system, but they must not take priority over gaining new insights and skills using the new system.

Some project managers attempt to ensure regular BI use by creating a usage report or two that needs to be monitored by managers or supervisors. However, such a report cannot characterize actual use. Was the report’s data used in making decisions or was it simply opened but ignored? Did workers retreat to the old report and incorporate its data in daily operations? This scenario frustrates DW consultants because “passive” consumption of BI solutions can be a harbinger of program failure. Imagine that following a seemingly successful parallel run (as demonstrated by usage reports), the project lead declares the old reporting system can be shut down. Users are suddenly left with conflicting reports that they failed to scrutinize properly during the parallel run.

A far more successful approach is to retire your old reporting framework upon implementation of the new. Use only one reporting framework and one source of data for all reports: tabular, printer-friendly, explorative, analytical, and so on. Your BI/DW framework should leverage all enterprise reporting needs. The following sections describe the essential components for achieving long-term buy-in and illustrate how significant adoption rates can make a BI project a source of excitement for the whole organization.

**Strategy #2: Enable Drill-to-Detail**

Although parallel runs pose problems for program success, practitioners must not introduce a new BI framework without providing users a means to verify its output. Project leads cannot expect users to adopt BI unless they are able to verify the new platform’s reports over a few reporting cycles.

An alternative for parallel run—that is, a pervasive mechanism to validate data—must be introduced. The best way to accelerate user confidence on a new reporting platform is to provide the capability to drill to detail (transactions) behind the summarized reports, metrics, and measures. This way, if a user feels a given number is not what is expected, he or she can verify it immediately from the same screen and gain confidence in interacting with the BI platform.

This technique prevents users from getting stuck matching data between old and new reports. The detail (drill-down) report must provide the proper information granularity. Usually, an individual detail report is
Developed for each business function, such as item-level detail for procurement, order-level detail for sales, and payment-level detail for revenue. Do not create a detail report for each summarized report because it can potentially recreate the reconciliation problem of parallel runs.

In addition to providing a better testing and verification mechanism, drill-down also serves as an integrated and single information delivery mechanism. Rather than having completely disconnected reporting frameworks for the summarized (aggregated) and detailed reports, drill-down provides an integrated platform that supports a robust and flexible architecture and makes data comparison easier.

Strategy #3: Integrate Data Definitions (Metadata) with BI
Drilling down to the transaction level is the first line of defense against critics of the new BI platform. What can make them feel empowered by the BI solution is ensuring that metadata is integrated into the reporting framework so that every piece of information on a report has its business definition attached as an interactive option (or in the form of a legend, depending on the type of report and technology used).

The usefulness of metadata is highly underrated. The metadata layer is essential for proper documentation of the technical aspects of a BI/DW project and equally important (if not essential) for helping both users and developers understand business measures and metrics and the rules behind them. Most users are concerned about BI usage because they get lost in the application. However, effectively placing the metadata can mitigate the risk associated with the user interface (and users’ opinions about the BI solution).

Another reason for users’ frustration with BI is that they don’t know what a measure or metric stands for in business terms. Providing the definition of the metric or measure to users in a quick, easy, and meaningful way in the report makes users more confident.

Definitions and descriptions for data, measures, metrics, and KPIs should be stored and integrated in a way that can be viewed as part of the same information delivery mechanism that delivers the reports. There are several ways to implement this integrated concept, starting from a straightforward glossary document published as part of the report page to a hover-over (pop-up) tip and hyperlinked metric name(s) that can be clicked to view definition, description, and calculation information.

Consider a KPI example, “AR Days.” This KPI is used in financial and revenue collection reports in most organizations. It can have two types of definitions: temporal and nontemporal. The temporal calculation freezes the view of the KPI state at a point in time—for example, at the end of each month. The nontemporal version will recalculate the KPI every time the report runs, providing the most up-to-date view. Suppose the old report on AR Days implemented a temporal definition and the new BI system uses the nontemporal version (to keep in line with its data warehousing objectives). A user accustomed to the old report will see that the AR Days value is very different in the new report and may conclude that something is wrong (triggering a call to the help desk, perhaps) or assume that the new BI calculation is incorrect (which could undermine a user’s confidence in the new data).

If metadata (data definition and calculation formula) was instead accessible from the same report window, the user could access the metric’s definition and calculation quickly and easily, preventing a call to the help desk and preserving his or her confidence in BI.

An integrated metadata repository is effective only if it is well maintained. There must be an application interface that selected users (usually data governors) can use to periodically update the definitions of the data elements used in different reports. The challenge here is to balance keeping the data dictionary current with managing security rights to control and update it.

Strategy #4: Embrace Action-Integrated BI
The previous three strategies for increasing BI use usually prove to be sufficient for a conventional business environ-
ment. This fourth strategy provides a capstone solution that is appropriate for most organizations, making your BI implementation an outstanding one.

Users tend to lose interest in the capabilities of a good BI framework and end up using it as a passive report infrastructure. To invigorate a few users and build excitement about using BI to its full potential, users must be able to act on the information in what we call an integrated “reporting and action” environment. This concept will soon be a focus of many BI installations.

Action-integrated BI is different from action-oriented BI. Action-oriented BI emphasizes the accuracy, measurability, and direct or indirect mapping and association of a KPI with the outcome of a business function. Action-integrated BI focuses on the association of actions with the performance outcomes of a KPI. In action-integrated BI, the actions are defined for each state (good, par, and below par) of a KPI. This concept is relatively new and few off-the-shelf BI products provide this functionality, but it may be an important feature in the next cycle of BI products.

Users tend to lose interest in the capabilities of a good BI framework and end up using it as a passive report infrastructure.

The action-integrated BI idea is simple yet powerful. In a typical scorecard or dashboard, KPIs are displayed with numerical values and graphical indicators with color codes (green, yellow, red, etc.) or up and down arrows depicting comparative performance against some preset benchmark. In action-integrated BI, another object is paired with KPIs in the form of a button or link that allows users to perform actions based on the indicator. This distinction means more than sending an e-mail message with the KPI to a co-worker; it means taking action—performing some business task—by sending business directive(s), such as “Keep stores open late for the season,” to managers in a particular region.

Such functionality requires more effort at the business end than on the technical side. This feature would initially seem to require additional business analytics reporting functionality combined with workflow engine functionality, but there are key differences. The only workflow functionality used in the action-integrated BI solution is the ability to send e-mail messages to, or initiate actions by, different users (if you consider actions as work items), which is only a very small part of a truly good workflow engine.

At a technical level, what is required in an action-integrated BI solution is the ability to assign multiple actions or better action categories as part of the KPI definition. A new and separate database application is required to define and maintain a list of actions. The application should be designed in a way that allows multiple levels of actions to be created and associated with action categories, action types, actions, subactions, and so on. The reporting infrastructure must be able to represent the KPIs with KPI indicators and provide an action link or button for each KPI. The action link will open an applet that allows users to select actions defined for that KPI.

In a conventional BI environment, the business unit is usually responsible for providing and maintaining KPIs for an organization, which can be a challenge. For an action-integrated BI solution, the business unit will be required to define and maintain the list of actions and action types tied to KPIs in addition to the underlying KPI definitions. There is a risk of poor list maintenance where the actions become irrelevant for the KPIs, and you must mitigate this risk. This level of interactivity and operational efficiency is a futuristic feature that should be attempted only after ensuring successful implementation of the initial three strategies described.

**Security Considerations**

Apart from the features that directly interact with users, there are also back-end factors that contribute to the successful implementation of a BI project in an organiza-
One such factor is the creation of a detailed data access policy and robust security model to support the data access restrictions in a BI framework.

The need for, and significance of, a good security model versus an on-the-fly security model is not realized if the number of BI users is small. After the initial deployment, upper management usually likes the new technology, hails IT's effort, and immediately wants to provide BI access to their teams. The catch, though, is that senior executives want their teams to use BI but with limited data visibility to avoid confrontation among team members. Your BI solution needs a robust security model designed from the beginning of your project to support a wide variety of data security permissions. Without this, managers may not approve your project implementation.

A robust security model must be able to restrict access:

- At the data level
- At the metric level
- At the BI component level (dashboard, scorecard, report)

The need for, and significance of, a good security model versus an on-the-fly security model is not realized if the number of BI users is small.

Quite often, IT stumbles into implementing the security policy at different levels rather than creating a centralized user management application that is responsible for creating and maintaining user profiles and access to BI and all other applications in the organization. This is especially useful for organizations that start their BI implementations on a modest scale but expand rapidly. In such cases, user management becomes a hassle for IT support and restricts enabling BI usage across the organization.

Summary

Successful implementation of a BI project depends on the readiness of the end users to try out the new reporting framework. The strategies described in this article will help you prepare your environment so that end users can easily access the information without getting lost in the BI platform's traps, such as data reconciliation or not understanding basic calculations. These real-world, tested, and proven strategies will enable project managers and BI owners to create an environment that enables the whole organization to use BI with all of its powerful new features, gain user acceptance, and increase adoption. That is what pervasive BI is all about.
BI Case Study

Apparel Company App Melds Fashion, Mobile BI

Using iPads in the field, buyers can make immediate decisions that impact sales and revenue.

Linda L. Briggs

The apparel company Guess?, Inc. is known for cutting-edge fashion, so it stands to reason that the firm is making business intelligence (BI) look good, too. Using Apple iPads and mobile BI software from MicroStrategy, the worldwide clothing chain has created a BI app that gives Guess buyers and executives vital information at any time, whether in the office or at one of the company’s 1,465-plus stores worldwide. The app makes BI accessible to nontraditional users by presenting data in a well-styled, easy-to-understand package whose multiple screens say “fashion” long before they say “data.”

Guess designs and distributes apparel and accessories such as clothing, handbags, watches, and shoes. The company emphasizes innovative and distinct product designs and uses a team of designers who seek global fashion trends for the company’s style-conscious consumers. The Guess approach is working—despite 2009’s worldwide economic slump, the company increased global sales by 20 percent to $2.1 billion. It also expanded into key markets worldwide, including opening new store locations in the U.S., Europe, and Asia.

Going Mobile Early

Guess started supporting its mobile workforce users in 2008, when it began to deliver critical information to RIM BlackBerry devices belonging to executives and corporate directors at headquarters, along with regional directors in the field. In 2009, the company added dashboards to its mobile BI offerings, using stylish, well-designed screens to provide information to nontraditional BI users such as buyers and planners.

In 2010, director of BI Bruce Yen and his group at Guess took the mobile initiative a giant step further and rolled out a cutting-edge MicroStrategy app for the iPad. The spark was Yen’s attendance at a MicroStrategy conference—the company uses other MicroStrategy products—where he realized that the BI company had a workable mobile app ready for the iPad (a device Yen admits he had previously been skeptical of).

A MicroStrategy presentation, however, “really piqued our interest,” and in brainstorming with Guess’s CIO, Yen says, “We thought, maybe there are some business issues out there that we can use this device to tackle. … We
settled on a subject area and a group of users that we thought could really use the app.” That group was Guess buyers, who spend large amounts of time traveling to retail stores and meeting with store managers as they make purchasing decisions.

As the idea developed further, and as MicroStrategy’s mobile BI offering matured, a proof of concept followed, along with much discussion, experimentation, and development work in an area that was still quickly evolving. As work continued, the pioneering nature of the project was clear—at one point, Yen and his team submitted more than 50 issues and enhancements to MicroStrategy. The app itself was initially released to pilot users in November 2010 and went live in the spring of 2011.

The iPad app replaces Microsoft Excel spreadsheets and other methods that buyers traditionally used for reviewing data—many of them simply too cumbersome to be used in the field, or even at the office. The app, in contrast, delivers quick and visually appealing operational snapshots of the business. Using that information, buyers can make the sort of immediate decisions that impact sales and revenue. Most of the data is pulled from the company’s retail data warehouse, Yen says; some added work was done to import other data, to build additional metrics and analytics, and to add special features such as geolocation coding that shows the exact latitude and longitude of stores.

Key to Yen’s strategy for the iPad mobile BI project was the idea of an actual app, rather than a set of dashboards that would simply display noninteractive information to users without any sort of consistent tie. In its current iteration, the multifaceted app actually combines the equivalent of some 12 dashboards into one, so it can be used not only by buyers but by executives and buyers at home or at the office as well.

Another key to the project was keeping the technology aspects well removed from users. The term “business intelligence” simply wasn’t used with the target audience, Yen says. “We didn’t market it as a BI app; we called it the Gmobile app,” he explains, in the hope that users would be naturally drawn into using it.

“Every pixel was looked at on every single slide or dashboard we built.”

To that end, his team also tried to avoid “giving users the feeling that they were going into a world of dashboards and data.” Instead, he worked to design an app that gives users an enticing starting point—an opportunity to be briefed on the business at any one point in time, essentially, and then to dig deeper if they choose.

Ahead of the Infrastructure Curve

The initial thought was that buyers would use the app wirelessly at Guess stores, but the project proved to be ahead of the infrastructure—some stores didn’t have the wireless capacity to support the iPad. Instead, traveling business users draw on hot spots or use the wireless capacity offered by hotels.

In fact, users have surprised Yen and his team in the ways they use the app. Rather than employing it mostly at stores, as the team envisioned, they often jump into the app before they reach a store, to brush up on store management and sales history. “Many users told us, ‘It’s a great app, and I don’t always need it at the store. I use it more before I get to the store.’” Users also rely on the app in their offices and in meetings, often preferring it to running a report, since they can grab detailed information quickly and precisely with a few taps.

The popularity and ubiquitous nature of the iPad meant that users wanted to run the app on their personal iPads. This practice brought up security issues, as did the app in general, because it essentially allows access (albeit password protected) to confidential store data from anywhere. To increase security, along with passwords and data encryption, Yen uses third-party software for mobile device management, and can shut down access to a device if it is reported stolen, for
example. Allowing personal devices to access corporate data brought up the pioneering nature of the project yet again, he says, as MicroStrategy continued to fine-tune its approach to security in its mobile BI offering: “The issue has long since been solved, but at the time, we had to figure out how we could [allow personal iPads to be used] and still protect our corporate assets.”

**A Critical Component: Design**

Perhaps not surprisingly for a company whose brand revolves around fashion, Yen says that the appearance of the app was critical. He worked with a graphic artist and emphasizes how crucial the look and feel is. “I don’t think many BI teams think design is that important,” he says, “but I think it is….Every pixel was looked at on every single slide or dashboard we built.” That scrutiny included colors, images, and the background itself. To test out designs, Yen invited co-workers with a good fashion and design sense—and a good feel for the Guess image—from different departments for impromptu feedback sessions. “It wasn’t anything formal—I just called people and said, ‘Hey, do you have 10 minutes?’”

Because the app presents so much data from so many angles and can drill to varying depths, it can serve a range of user types, from executives to buyers to designers. By tapping on the screen, users drill down into new screens with additional data. The iPad’s famed instant-on capacity, long battery life, sharp display, and ease of use help, making the software accessible to nontraditional users and helping transform how they view BI.

In fact, the iPad’s intuitive feel has meant little or no user training. Yen says user feedback was incorporated along the way to make the system as intuitive as possible, but no formal training has been requested. To ensure usability as the app evolved, he traveled with buyers as they visited retail stores, watching for any problems such as performance and connectivity issues in the field. As the project continues to evolve, he says, the balance between ease of use and intuitive design, while still presenting plenty of solid business information, is largely there. “All in all,” Yen says, “I think it’s intuitive enough for someone to just pick it up, use it, and not have a lot of questions.”

In summary, Yen says he sees content as key, but it must be offered in formats that appeal to the intended audience and emphasize ease of use. “Usability and workflow were two of my main focuses with this, along with, of course, the visual aspect,” he says. Because Apple’s designers have set the bar so high for iPad apps with the product’s design, he advises companies planning to launch an iPad app to budget a bit extra on design. As the Guess app has shown, the payback in pervasive BI—although users will probably never call it that—will make the investment more than worthwhile.
Editorial Calendar and Instructions for Authors

The Business Intelligence Journal is a quarterly journal that focuses on all aspects of data warehousing and business intelligence. It serves the needs of researchers and practitioners in this important field by publishing surveys of current practices, opinion pieces, conceptual frameworks, case studies that describe innovative practices or provide important insights, tutorials, technology discussions, and annotated bibliographies. The Journal publishes educational articles that do not market, advertise, or promote one particular product or company.

Editorial Topics for 2012
Journal authors are encouraged to submit articles of interest to business intelligence and data warehousing professionals, including the following timely topics:

- Agile BI
- Architecture and deployment (including cloud computing, software-as-a-service)
- BI adoption and use
- BI and big data
- Data analysis and delivery
- Data design and integration
- Data management: MDM, data quality, and data governance
- Data warehouse and database technologies
- Mobile BI
- Project management and planning
- Selling and justifying the data warehouse

Editorial Acceptance

- All articles are reviewed by the Journal’s editors before they are accepted for publication.
- The publisher will copyedit the final manuscript to conform to its standards of grammar, style, format, and length.
- Articles must not have been published previously either online or in printed form. Submission of a manuscript implies the authors’ assurance that the same work has not been, will not be, and is not currently submitted elsewhere.
- Authors will be required to sign a release form before the article is published; this agreement is available upon request (contact journal@tdwi.org).
- The Journal will not publish articles that market, advertise, or promote one particular product or company.

Submissions
For more information and complete submissions guidelines, please visit tdwi.org/journalsubmissions.

Materials should be submitted to:
Jennifer Agee, Managing Editor
E-mail: journal@tdwi.org

Upcoming Submissions Deadlines
Volume 17, Number 2
Submission Deadline: February 24, 2012
Distribution: June 2012

Volume 17, Number 3
Submission Deadline: May 18, 2012
Distribution: September 2012
Making Your Organization Care about Data Quality

Mario Cantin, b.Eng., MBA, PMP, MCDBA, MCSE, has been a data quality strategist for over 20 years. He is an independent consultant specializing in data governance, data quality management, and data integration strategy. mario.cantin@thedataqualityframework.com

Abstract
Data quality (DQ) is a tough concept to grasp. We know that data quality is important, but what does it mean to manage data quality? What are the pragmatic drivers for ensuring an organization cares about the quality of its data?

Organizations are driven by establishing goals and realizing benefits. Managing data quality is managing the links between quality of data and business value. It’s about defining and ensuring, from a data perspective, the accountability of the organization toward the processes.

Projects that implement data processes present an extraordinary opportunity to improve data quality by clearly linking quality of data to project benefits. This article presents a simple, pragmatic approach to define and ensure project accountability toward data quality and organization accountability toward data processes. The approach can lead your organization to care about data quality.

Introduction
We all care about data quality and understand that it is important. We are well aware that data is “the new asset.” We understand that in our global, competitive business world, taking the right path depends on the quality of the data used to make decisions. We assess data, we profile data, we govern data, and we master manage our data. We care when our efficiency, processes, objectives, and bonuses are negatively impacted by poor data quality.

Everyone cares about data quality—yet nobody cares. No one really cares about how the data we manage will be used by other sectors or other departments within our organizations. We don’t care because we usually have no idea how, why, and when our organization will use our data. We manage our data’s quality in line with our
processes, requirements, and objectives, rarely aware that this approach might affect other sectors of the organization. If we are aware of the issues, we do not devote time to them.

For example, the date of birth (DOB) of a client may not be mandatory in the billing system of an organization, and so its level of completeness may be only 60 percent. Billing probably does not consider its data to be of poor quality because, from its standpoint, the DOB is probably information that is merely “nice to have.” However, the marketing department views DOB as a key element for campaigns. Thus, from the marketing department’s perspective, the client data in the billing system is not high-quality data.

It’s All about Processes

The technical dimensions of data quality (including completeness, correctness, availability, and conformity) are attributes; they do not represent the actual data quality. What exactly is data quality? Data quality is the alignment between business data requirements and the level of completeness, accuracy, and availability of the data. An organization that is effectively managing data quality does not need to have 100 percent complete, 100 percent correct, and 100 percent available data. It simply needs to ensure it has the level of quality required by its business processes.

Data quality must be driven by business value, objectives, and priorities. Analyzing the completeness or uniqueness of data in a database is a technical task. It’s a snapshot that doesn’t provide any drivers to improve the quality. Caring about data quality means ensuring that the data consumed by processes adequately supports the process objectives. Why care about the quality of unused or poor-quality data if it doesn’t impact strategic or operational objectives? Who cares about improving the quality of data when it isn’t clearly linked with business value?

The recent significant interest in data quality stems mainly from process integration. By integrating processes (and thus integrating data), we are basically defining new data quality requirements. Data that was perfectly suitable for legacy processes suddenly becomes “incomplete” or “incorrect” when used in a new process. Data issues are raised—though the issues are not really about the quality of the data but rather about the impacts of the quality of that data on the new processes.

Organizations must establish, implement, and follow good data management practices through data governance, data architecture, database design and administration, and security. IT governance has the critical role of ensuring that the organization uses proper data standards, methods, technologies, design, and tools. To improve the quality of data and promote organizational cultural change toward data quality, data management must be aligned with business drivers. The management of data quality must be driven by a regard for the data’s impact on business processes. Where the lack of quality data puts the output of a process at risk, the organization must make improvements.

A lack of quality data in a database or a system may not directly drive an organization to “clean up” its data, but when a process does not deliver expected benefits because the data it uses is incomplete or incorrect, this will drive such an effort. By clearly linking poor quality to negative effects on business processes, you will implicitly build the business case to improve data quality.

The real driver for data quality improvement is the failure of processes to deliver their expected benefits if data is incomplete or incorrect.

If data quality is about impacts to processes, we’re faced with several questions. How can we make sure the quality of data is in line with process requirements? How can we promote an organization’s cultural shift toward data quality? How do we change a strong paradigm centered on local data validation and segmented use of the data?
How can we make sure data creators manage their data to the standards the business requires and not simply to meet their own requirements?

To answer these questions, we must examine two foundations of data quality accountability (see Figure 1):

- Business processes that consume data are accountable for defining and communicating their data quality requirements.
- Data providers must be accountable for the quality of data they provide.

As simple and straightforward as they may sound, assigning and monitoring such responsibilities is the key challenge of managing data quality. One problem is that data quality is not a trivial, implicit request. Business users must understand how managing data quality ties technical considerations directly in with business value. Requesting data quality is like asking the organization to be efficient. What do you mean by efficient? What do you mean by data quality?

**Making Organizations Care**

Accountability is the key. The organization’s departments must be accountable for the quality of their data, not only as they use it, but also as the rest of the organization uses it.

Data providers cannot simply be told that they need to provide good-quality data; they must be told what data is important and why. For example, the billing department cannot just be told that data quality is important. Because billing department staff members do not consider incomplete DOB in client files to be a quality issue, they need to know that DOB is key data for the marketing department.

An organization must define, quantify, and communicate expectations in terms of data quality for all processes. Data quality metrics for key data processes are effective accountability drivers. As we’ll explain, metrics for key data quantify data quality and allow integrated corporate processes to clearly define and communicate critical data elements that require attention and focus.

Projects that integrate data are accountable for the quality of the data they deliver. In projects integrating data from different sources, the data quality of the source systems is usually viewed as an uncontrollable constraint with which the project must deal. Project managers do not want the success of their project to rely on factors that cannot be planned in a detailed manner, so data quality management is rarely part of project objectives and deliverables.

Yet, projects are approved based on business cases and anticipated benefits, so data integration projects are accountable to manage all the risks that might affect their business case. They need to identify, measure,
and evaluate the impact of poor data quality on the processes they’re implementing. Recommendations and plans to mitigate data quality risks must be defined and communicated.

Projects cannot solve data quality problems for the entire organization, but they need to clearly identify how these problems will impact their business case and must offer solutions. A project cannot be successful if, once in production, the data that flows from it is not in line with expected outputs. To deliver data quality, a project must link its success to it, not dissociate from it.

**Opportunities to Improve Data Quality**

Data integration projects such as customer relationship management (CRM), master data management (MDM), and business intelligence (BI) present a unique opportunity to improve data quality. These projects define and implement new data processes. Through their business cases, business value and anticipated benefits are associated with each process. When data quality is evaluated in terms of its impact on business processes, data integration projects present a great opportunity to clearly link data quality with business value.

A project cannot be successful if, once in production, the data that flows from it is not in line with expected outputs.

Managing data quality within projects is not easy. The first challenge is to convince the project managers to manage data quality. Because data quality management is usually perceived as a good practice rather than a criterion of project success, the common response of project managers is to position data quality as an architecture and development concept. However, architecture and design are not responsible for ensuring that the data flowing in from other systems will support implemented processes.

These fundamental practices will ensure that data quality becomes a priority for the project team:

- **Define** the delivery of quality data as a quantifiable success factor for the project
- **Detail** the data quality requirements in the business requirements documents
- **Define** a data quality stream within the project and assign a data quality lead; ensure he or she reports to the project manager or business sponsor
- **Define** key performance indicators (KPIs) to monitor the management of data quality during the project

**Making Project Managers Accountable for Data Quality**

Projects must deliver data quality! Project managers usually don’t like this. How can their project be accountable for something they perceive as being out of their control? It is much more convenient to say the project will deliver systems, functionalities, and processes, but how can a project control the quality of the source data that will flow from other systems? The quality of data from external systems is usually considered to be out of scope.

In the next sections, we’ll examine a pragmatic method that ensures data quality is delivered at the end of the project and that data providers have clearly defined and enforced accountabilities. Three main steps should be taken:

- **Identify and mitigate** any risks that poor-quality source data might impact the processes that are being implemented.
- **Define** key data for each process. Ensure that adequate data quality controls are implemented for key data that require attention and focus.
- **Ensure** clear organizational accountabilities by defining, implementing, and publishing data quality metrics for key data.
The remainder of this article is devoted to our proposal for a highly effective way to motivate an organization to improve data quality, one key process at a time. See Figure 2.

**Step 1: Manage Data Quality Risk**
The project must manage the risks that poor data quality might create for the processes being implemented. Analyzing the risks by targeted process rather than by data source allows a project team to implicitly build the business case needed for change. The risk is not that the data in system X is of poor quality; the risk is that the benefits associated with an implemented process might not be achieved because the quality of data in system X is poor. Addressing data quality by process provides decision makers with the arguments needed to address the problem. Analyzing the risks for all the processes implemented by a project also ensures that all data risks are covered (see Figure 3).

**Risk Identification**
Identifying data quality risk is an iterative process. It starts early in the analysis phase and ends at project completion. As soon as the main data sources are identified and high-level target processes are defined, data discovery workshops can help identify data quality risks that will benefit significantly from being addressed early in the project.

There is usually a strong resistance on the part of the project manager, who may not want to spend time identifying data risks early on. The standard approach is to wait until target processes are designed. However, many data quality risks can be identified (and mitigated) based solely on the data sources and high-level vision of the targeted processes.

Data discovery workshops held during the analysis and design phases with data quality experts, subject matter experts, and source data experts help establish the
baseline of data quality risks. Early in the project, the
data quality expert can challenge the intended use of
the data source by the target processes. The goal is not
to identify issues related to the current processes but to
identify the risk to targeted processes that current data
problems will affect.

As target processes are defined, new risks will emerge. As
architecture and design are detailed, the risks related to
data quality must be addressed. Risks are also identified
by analyzing the results of data profiling and by assessing
the impacts of measured poor quality on the targeted
processes.

The risks that are hardest to address come from project
data interdependencies and risks related to external data
sources. By being involved early and throughout the
process—including while defining business require-
ments and architecture elements, designing the business
processes, and profiling the data—the data quality expert
can ensure the proverbial dots stay connected from a data
quality perspective. Thus, many risks are not identified
by specific project streams but rather through the data
quality project functions.

A data quality risk matrix can be used to summarize the
risk elements for each process. This matrix should show
the value associated with the process or the criticality of
the process as it relates to the overall project objectives.

These elements become the triggering and weighting
factors used to address the risks.

Risk Analysis
Data quality concerns and risks are raised during the data
discovery workshops and data profiling activities. Before
initiating any mitigating actions, you must analyze the
risks and understand the potential impacts (see Table 1),
which will build the business case to support your
recommendations.

Risk Mitigation
There are several options for mitigating data quality risks.

Accept the risk.
Accepting the risk might be the right option if the data
at risk represents only a small volume, if the impacts on
the process are minimal, or if the correction of the poor
data quality would require unjustifiable effort. Periodic
validations should be done throughout the project to
reassess the risk status.

Improve data quality in the source systems.
The business cases might recommend improving the
quality of data in source systems. This improvement effort
often falls outside the project scope, but the business case
may allow escalation of the issue to the executive level
and a clear demonstration of the link between poor data
quality and project benefits. Should the business improve
the source systems’ level of data quality or should the

<table>
<thead>
<tr>
<th>Data quality assessment</th>
</tr>
</thead>
<tbody>
<tr>
<td>• Is it truly a risk to the project?</td>
</tr>
<tr>
<td>• What is the volume of problematic data?</td>
</tr>
<tr>
<td>• How many records are at stake? How many clients? How many files? Are they active?</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Impact analysis</th>
</tr>
</thead>
<tbody>
<tr>
<td>• Which target processes will be impacted if the project deploys without addressing the risk?</td>
</tr>
<tr>
<td>• What are the values and/or benefits of the impacted processes?</td>
</tr>
<tr>
<td>• For each impacted process, when will the risk materialize (number of days before design, tests, or deployment)?</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Options and recommendations</th>
</tr>
</thead>
<tbody>
<tr>
<td>• What is the root cause of the poor quality?</td>
</tr>
<tr>
<td>• What are the options to mitigate the risk?</td>
</tr>
<tr>
<td>• What are the recommendations? The next steps?</td>
</tr>
</tbody>
</table>

Table 1. Data quality risk analysis and business case.
Data quality improvement projects can aim at doing a one-shot cleanup of the data or a cleanup over a defined time period. The cleanup usually comes with process improvements and changes in the systems themselves to reinforce data quality controls. For example, a project implementing a new application to manage marketing campaigns could highlight, in a business case, the potential impacts of poor-quality client data. A portion of the anticipated benefits of the marketing systems could be at risk because the clients’ DOBs, managed in the billing system, are only 60 percent complete. The identified cause might be that the DOB written in the client subscription forms is not systematically entered in the billing system.

The recommendation could be to hire a temporary resource to manually review the paper file and enter the information in the billing system; to adjust the billing data entry process to include the DOB; or to modify the billing system to make DOB mandatory when adding a new client to the system. The recommendation could also include additional quality controls in the billing system (such as cross-validation with a driver’s licence) to validate the DOB.

Improve data quality during ETL.
Data quality can be improved during extract, transform, and load (ETL). Data can be cleansed to remove the DOB of 700-year-old clients or of clients who died before they were born. ETL processes could also enrich incomplete data with information from other sources. Recommendations of this nature are often considered as a change request in the project.

Improve data quality in the target systems after deployment.
The recommendation may be to transfer the data with the current quality level and improve data quality in the target system, especially if the source system is to be decommissioned and the project timeline allows addressing the risk at a later time.

A Final Word about Risk Mitigation
When addressing data quality risks, the key success factor (and challenge) is flexibility. As new potential risks are identified throughout the project, priorities must be revised. You may have to delay elaborating a business case to assess another potential risk; you may have to drop a business case in its early stages if you discover the risk is not material. Risk identification and mitigation are not documentation activities. Data quality risk management must lead to measurable added value.

The risk matrix should be reviewed frequently by the project management team. Activities in the mitigation plan that are outside the project should be escalated to the corporate data steering committee, which should act as the bridge between the projects and the different sectors or lines of business (LOBs) of the organization.

Step 2: Implement Data Quality Metrics for Key Data
The project ensures data quality by identifying and mitigating data quality risks. Implementing data quality metrics (see Figure 4) will ensure that the organization
stays focused and accountable for delivering data with the level of quality appropriate for the new processes. To create the metrics, the project team must identify key data.

Key Data
All data is not created equal. When different data elements are incomplete or incorrect, they will have completely distinct impacts on the processes that use the data. Key data elements require accountabilities, action plans, and organizational attention.

Requesting that all data be 100 percent complete or 100 percent correct can rarely be justified by business value. New data quality requirements are driven by new implemented processes. The project must identify the data elements that are critical to the new processes.

Key data have the following characteristics:

- Their lack of quality can remain undetected throughout the data processes but impacts the process outputs.
- Their quality level requires business attention.
- Their level of quality can be measured.
- Their number is limited (the list is short enough for the business to act on); identifying measurable, manageable key data in this way makes the difference between simply talking about data quality and empowering users to improve data quality.

Step 3: Implement Data Quality Metrics
The project is accountable for identifying the key data for the processes it implements. Based on the key data, the project team must define and implement data quality metrics that will support data quality governance when the project is complete.

There are two types of data quality metrics: strategic and operational.

Requesting that all data be 100 percent complete or 100 percent correct can rarely be justified by business value.

Create strategic quality metrics on key data to define and enforce accountabilities.

The strategic data quality metrics act as the bridge between the process requirements and the business. By defining and implementing metrics for key data, the project defines the accountabilities of the organization toward the new processes. It’s the way to say to the rest of the organization, “Here’s what’s important to us.” Strategic data quality metrics are not used to discover data quality issues; their purpose is to ensure accountabilities and action plans about known issues.

During the project, the metrics should be defined in collaboration with data stewards in the different sectors of the organization. If data stewardship has not been defined, this is a good time to do so. Data stewards are accountable for the quality of data in their LOBs, so the project must inform them of their new accountability with regard to the new processes being implemented.

Collaboration with data stewards is key. The metrics defined must have meaning for the data stewards who are accountable for the data behind them. They should be able to act on it. A metric for “Completeness of DOB of client in the billing system” might make absolutely no business sense and thus departments will never be accountable for it. A better metric definition might be “Completeness of DOB of an active client in the billing system.” The data stewards will own the metrics, so they must be fully involved in their definition and design.

The definition of strategic data quality metrics is a deliverable of the project’s data quality team. The metrics
can usually be produced and published monthly in a data quality dashboard.

**Define operational data quality metrics to monitor and control the state of the data.**

These metrics are not limited to key data. They provide a window into the state of the quality of data currently used by a process, such as number of rejected records, number of unmatched clients, and alerts and thresholds, which are common operational data quality metrics. They are used for capacity planning and drive operational actions. Their analysis over time may also provide insight on strategic data quality issues.

During the project, the definition and implementation of these metrics is the responsibility of the architecture and design team. These metrics should be detailed in the business requirements, as they are known and required by operations. After project delivery, operations own the metrics and are accountable to act on them.

**Dimensions of Data Quality**

Data quality metrics measure dimensions of the data quality. Two of the key dimensions used for strategic management of data quality are completeness and correctness.

**Completeness**

Does the data exist? Completeness is often the easiest quality dimension to measure. Data profiling tools or simple queries in databases can provide the metric (e.g., “DOB completeness: Percentage of active clients with a valid DOB in the billing system”).

**Correctness**

Are the values of the data correct? What is our level of confidence in these values? Defining correctness metrics usually requires creativity. How do you define a metric that will tell you every month the level of confidence in the correctness of the DOBs in the billing system? Here are a few suggestions:

- **Sampling and replication:** Periodically extract a sample of the client records and validate the DOB with an external source. For complex values such as credit scores, manual recalculation of the values in the sampled files will indicate the level of confidence.

- **Operational checkpoint:** For example, obtain from the insurance department their operational metrics showing how many of their insurance proposals had to be reviewed in the last month due to incorrect DOB.

- **Reliability indicators:** The fact that the client’s information has been validated with his ID may reinforce the level of trust in the DOB in the system. Correctness metrics could be derived from the percentage of clients that have an ID number in the system.

**Data quality metrics measure dimensions of the data quality. Two of the key dimensions used for strategic management of data quality are completeness and correctness.**

Other dimensions of data quality, such as availability, conformity, and security, are usually key drivers for architecture and design rather than strategic data quality management.

The implementation of data quality metrics is a challenge (as are all other aspects of data quality management). The strategic metrics, based on identified key data, are often defined late in the design phase of the project. At this point, you need architects and developers to define and implement a solution. The metrics can be implemented in the source systems, ETL process, data warehouse, or target systems. Some metrics may require manual processes (e.g., replication sampling), automated jobs, and technical extraction processes. Because the expertise required to implement the metrics is hard to plan for before metrics are defined, implementing strategic metrics is often handled as a subproject task or a change request.
Data Quality Dashboard

Publishing the data quality metrics in a corporate data quality dashboard is the true leverage of data quality metrics. The data quality dashboard is a strong accountability driver. Strategic data quality metrics are not used for monitoring data quality; they are used to reinforce accountabilities. A data quality dashboard presents:

- Key processes or risks on which the organization is focusing
- Data elements that are key to these processes and that require actions
- Business data stewards accountable for the metrics
- Metric value and target value
- Link to the action plan (or absence of action plan)

If your organization does not have a corporate data quality dashboard, integration projects present a good opportunity to define and implement one. A data quality metric repository can be designed to store the metrics and feed the dashboard. As new projects implement new processes and new metrics, they will benefit from the data quality repository and enrich the dashboard with new processes that are key to the organization.

Success Factors

Managing data quality in projects is a challenge, but the following key factors will ensure successful data quality management.

Position data quality under business management. A data quality risk is a business risk, not an IT risk. IT can meet its objectives without data quality, but the business cannot. Data quality management requires IT skills driven by business objectives and exposure.

Get upper-management buy-in. Accountabilities to deliver data quality will not emerge by themselves from the projects or the LOBs. Executive sponsorship is critical.

Start as early as possible. Data quality management should be involved at the very beginning of projects to ensure that data quality requirements are part of the business requirements, to position the data quality stream, to ensure adequate funding for data quality, and to identify and address data quality risks.

Include data quality in project KPIs. Projects are usually too busy meeting milestones to worry about data quality. During design sprints, getting the project manager’s attention to address risks that data may not support the processes is a challenge. Define KPIs that will highlight the efficiency of the project to manage the data quality risks (e.g., number of data quality risks identified, number of business cases submitted, number of business cases approved, etc.).

Communicate a sense of urgency. When not managed proactively, data quality issues usually arise during the “build” phase of projects. At this point, they impact the project. The urgency to identify and address data quality risks in the early phases of a project will face resistance. The most common objection is, “How can the project address data quality risks before target processes are well defined and designed?” The main data quality risks are not related to detailed architecture and design. If these risks are not addressed and mitigated early in the project, they will impact the project.

Leverage the use of data profiling tools. Data profiling tools are often used only to analyze data from a data transfer or database design perspective. Use the tools from a strategic perspective to produce recurrent metrics for key data.

Define and communicate the data quality management plan. Explain the vision, method, and deliverables to the project team. Get their buy-in.
Enabling BI Agility with Data Virtualization

The need to get information to the right people at the right time is what business intelligence (BI) is all about. Unfortunately, IT’s backlog often makes it impossible for an enterprise to keep ahead of its competition, make smart business decisions, or get a full and timely look at a single version of the truth. The case studies featured in Data Virtualization: Going Beyond Traditional Data Integration to Achieve Business Agility explain how data virtualization solved this problem for 10 large enterprises.

Many BI-related books I’ve read spend a great deal of time talking about theory, using vague, tech-laden terms that don’t help readers understand the basics. They go on and on about processes and dynamics and platforms, but it all sounds like hand-waving to me. Not so with Data Virtualization, which is divided into two parts. In the initial chapters, the authors provide a useful and down-to-earth introduction to the subject; it’s clear they have a keen grasp of the topic and can explain not only what data virtualization is but why you should care—without getting down and dirty into details that no BI practitioner needs to know.

The second part of the book consists of 10 case studies; Davis and Eve offer observations about what the case studies tell us about the technology and its use, along with the benefits enjoyed by each case study subject.

In addition to clarifying the fundamentals of data virtualization, the authors explain its basic benefits: it all boils down to a three-part focus on agility. Davis and Eve reveal how data virtualization provides business-decision agility (knowing what has happened and what is happening so you can “identify and execute the ongoing business changes required for long-term success”), time-to-solution agility (reducing the time between when a need is recognized and when IT implements a solution), and resource agility (maximizing the business impact of and return on IT investments).

The authors contrast data virtualization with traditional data integration and the latter’s limitations and problems in today’s IT environment. They explain time and resource constraints: extract, transform, and load (ETL) processes and data replication take time, use additional storage space, and can introduce data errors; architectural complexity makes development difficult; and the frequently changing requirements IT must deal with aren’t compatible with traditional software development approaches.
After delving into how data virtualization technology works (discussing such elements as integrated development environments, server environments, and platforms), the authors turn their attention to how enterprises typically get started with data virtualization. The “Overview” chapter is particularly useful in this regard. The authors detail how data virtualization adoption goes through many of the standard application development “life cycle” steps—from business justification to technical evaluation to adoption.

Davis and Eve explore five common usage patterns that show the versatility of the technology, from BI data federation to an enterprise data virtualization layer, so you’ll get a better sense of what the technology can do and whether it’s suitable for solving your own data integration problems.

In the chapter titled “How Data Virtualization Delivers Business Agility,” the authors finally get down to the meat of their central thesis: to meet today’s “dynamic, complex, and challenging” business environment, businesses need to run their business and meet new challenges, yet they’re faced with problems ranging from data complexity to query performance. Businesses need complete, high-quality information and they need it now to stay ahead of their competition. They need to be agile. Data virtualization provides exactly that agility.

The authors list a host of impressive benefits. Data virtualization “greatly reduces complexity and reduces or eliminates the need for data replication and data movement.” Without the need for physical data movement, IT eliminates concerns about processing windows and capital expenditures for storage hardware (and backup protection for this storage, which the authors don’t mention).

With fewer moving parts, an enterprise can reduce “the need for coordination and synchronization activities.” Data virtualization enables an iterative development process and the ability to “fail fast,” a hallmark of agile development. Solutions are easier to develop, and IT can change them more easily (especially important with ever-changing requirements).

In fact, once the solutions are implemented, changeability comes into play again—data virtualization enables data sources to be easily added, removed, or replaced. A Fortune 50 financial services firm was able to change data sources “without requiring any changes to the consuming applications.”

Thanks to graphical integrated development environments, IT can play “what if” and build prototypes more quickly, and end users can see results and refine requirements faster, not to mention enjoy automatically generated code and built-in query optimization as bonuses.

Learning from Case Studies
Davis and Eve have selected “ten real-world case studies that demonstrate the significant value and tangible business agility benefits that can be achieved through the implementation of data virtualization solutions.” Each case study includes a brief overview of the company itself (four study subjects are not named), a description of the business problem to be solved, details of the solution developed, an overview of the implementation process, a brief discussion of the benefits and return on investment (ROI), a short list of critical success factors, and perspective on “future directions” (the company plans for applying data virtualization to other problems).

There’s plenty to learn, especially when the authors reveal some unexpected and less-well-known benefits. For example, Comcast found that accessing data virtualization technology (specifically in Composite Software’s solution) enabled access to LDAP databases from SQL, which increased its “ability to make additional use of the information.”

The ability to maximize use of existing data using data virtualization was a common theme among case study subjects. Compassion International, the not-for-profit child development and sponsorship...
organization profiled, found that the “data virtualization enables all data sources to act as a single source, making it easier to deliver data in multiple ways for different consumers.” A financial services provider explains how it could take its Web-accessible billing system and make it “available through an enterprise portal or as an app on a mobile device.”

Enterprises also benefited from better performance; for a Fortune 50 computer manufacturer, data virtualization eliminated the data latency and redundancy problems of ETL processes.

Despite the claim that the case study organizations “represent a diverse set of industries,” four of the 10 are in the financial services sector, and a fifth (NYSE Euronext) is a collection of equities and derivatives exchanges. Furthermore, all are large enterprises (one is a nonprofit, while the others are for-profit corporations) that chose solutions from Composite Software, where Eve is executive vice president of marketing.

The selection is understandable; from personal experience, I know that it’s hard to find companies that are willing to speak on the record about their problems (one vendor told me they thought it showed the public that the company was “weak”). The notion that disclosing one’s success exposes a firm’s competitive advantage (as another vendor explained to me) has always struck me as laughable, but I’ve no doubt that’s what the authors were up against when they tried to find case study subjects.

Although small and midsize businesses aren’t represented in the 10 case studies, they share many of the same problems, so the lessons learned are still valuable. Even if you aren’t an energy company or healthcare provider with 40 external data sources to integrate and a budget of $1 million (the problem faced in one case study), you’ll find most of the case studies interesting for their insight into how the technology can solve problems. Just don’t consider this set of case studies as a cross section of today’s IT challenges.

The book may disappoint some readers. Although they mention cloud data integration as one use case for data virtualization, the authors essentially ignore the cloud (it’s mentioned in only one of the case studies). The study of a “global 50 energy company” was too generic; after reading the author’s introductory chapters, this case study didn’t offer anything that hadn’t been discussed earlier in the book.

Also missing from most of the case studies are solid, definitive ROI calculations or project financials. For the computer manufacturer, “millions of dollars per year are saved through faster inventory turns,” but there’s no mention of how much the project cost. We have to take it on faith that “the company has successfully demonstrated significant ROI.” If you’re looking for cost/benefit analysis you can show your boss to gain support for your own project, look elsewhere.

One financial services firm acknowledged that selling the technology is tough, but “teaching them a new trick” can bring benefits. If you’re curious about those benefits (and there are many) or about whether the technology is right for you, Data Virtualization is a good place to start.

James E. Powell is editorial director for the Business Intelligence Journal. jpowell@tdwi.org
The State of Metadata Management

The Technology Survey that TDWI circulated at its World Conference in San Diego asked attendees to answer a few questions about metadata management as it is practiced in their organizations. TDWI ran the same survey in May 2010.

- **Metadata management serves many applications.** When asked which applications they manage metadata for, more than 80% of survey respondents pointed to business intelligence and data warehousing, far more than for any other application type (see Figure 1).

- **The spreadsheet continues to be the most commonly used tool in managing metadata.** That’s according to 30% of survey respondents in both 2010 and 2011 (see Figure 2). Even so, many organizations have availed themselves of software automation designed specifically for metadata management, as seen in custom solutions built in-house (23% / 13%) and dedicated metadata management tools (17% / 18%).

- **Metadata isn’t just about the data’s technical attributes anymore.** Traditionally, metadata has been literally “data about data,” describing technical attributes such as data type, data structure, and data sources or dependencies. More and more, metadata practices have expanded to describe other attributes, some of them business oriented, such as business definitions and data usage rules. Roughly half of users surveyed follow the traditional approach of describing technical attributes, whereas the other half also includes business-related information in metadata. TDWI feels confident that the trend toward stretching metadata’s uses will continue.

—Philip Russom, TDWI Research Director for Data Management

### For what kind of enterprise applications is your organization actively managing metadata? (Select all that apply.)

<table>
<thead>
<tr>
<th>Application</th>
<th>2010</th>
<th>2011</th>
</tr>
</thead>
<tbody>
<tr>
<td>Business intelligence and/or data warehousing</td>
<td>81%</td>
<td>86%</td>
</tr>
<tr>
<td>Data integration</td>
<td>44%</td>
<td>38%</td>
</tr>
<tr>
<td>Data quality</td>
<td>32%</td>
<td>32%</td>
</tr>
<tr>
<td>Data governance</td>
<td>29%</td>
<td>21%</td>
</tr>
<tr>
<td>Financial applications</td>
<td>28%</td>
<td>18%</td>
</tr>
<tr>
<td>Content management and/or Web content</td>
<td>27%</td>
<td>15%</td>
</tr>
<tr>
<td>Customer relationship management (CRM)</td>
<td>27%</td>
<td>19%</td>
</tr>
<tr>
<td>Customer data integration (CDI)</td>
<td>26%</td>
<td>18%</td>
</tr>
<tr>
<td>Master data management (MDM)</td>
<td>22%</td>
<td>16%</td>
</tr>
<tr>
<td>Data stewardship</td>
<td>19%</td>
<td>19%</td>
</tr>
<tr>
<td>Enterprise resource planning (ERP)</td>
<td>16%</td>
<td>6%</td>
</tr>
<tr>
<td>Human resources applications</td>
<td>16%</td>
<td>7%</td>
</tr>
</tbody>
</table>

**Figure 1.** Based on 462 responses from 124 respondents in 2010 and 284 responses from 95 respondents in 2011.

### What type of tool is most often used to manage metadata in your organization?

<table>
<thead>
<tr>
<th>Tool</th>
<th>2010</th>
<th>2011</th>
</tr>
</thead>
<tbody>
<tr>
<td>Spreadsheets</td>
<td>30%</td>
<td>30%</td>
</tr>
<tr>
<td>Custom solution built in-house</td>
<td>23%</td>
<td>13%</td>
</tr>
<tr>
<td>Dedicated metadata management tool</td>
<td>17%</td>
<td>18%</td>
</tr>
<tr>
<td>Metadata management facility within broader data management tool</td>
<td>13%</td>
<td>24%</td>
</tr>
<tr>
<td>Microsoft Access or other personal databases</td>
<td>5%</td>
<td>2%</td>
</tr>
<tr>
<td>Custom solution built by consultants</td>
<td>3%</td>
<td>4%</td>
</tr>
</tbody>
</table>

**Figure 2.** Based on 120 respondents in 2010, 91 in 2011.
Set yourself apart from the crowd. Get certified.

WHAT SETS YOU APART FROM THE CROWD?

Distinguishing yourself in your career can be a difficult task. Through TDWI’s CBIP (Certified Business Intelligence Professional) program, we help you define, establish, and set yourself apart professionally with a meaningful BI certification credential.

Become a Certified Business Intelligence Professional today! To find out what CBIP exams you should take, how to prepare, and where you can take the exams, visit tdwi.org/cbip.
TDWI Partners

These solution providers have joined TDWI as special Partners and share TDWI’s strong commitment to quality and content in education and knowledge transfer for business intelligence and data warehousing.