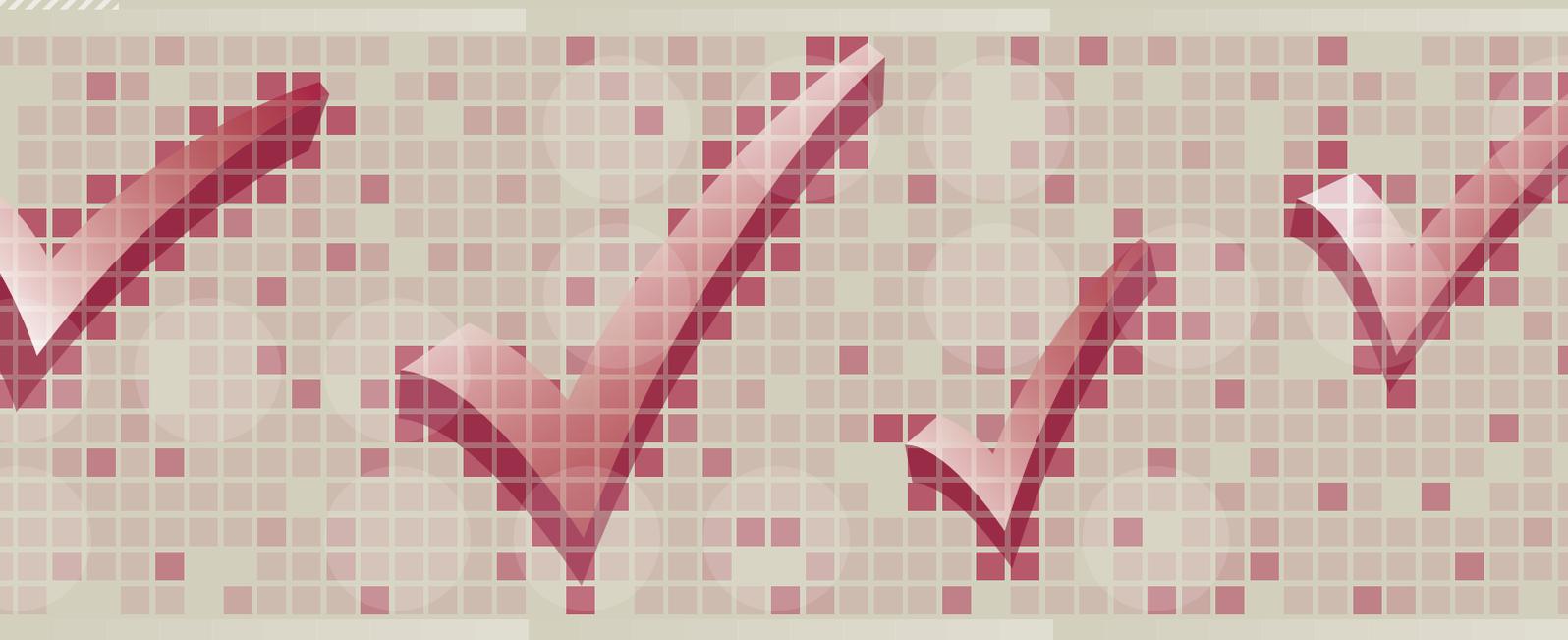


# TDWI CHECKLIST REPORT

Data Discovery and Unified Information Access:  
Expanding BI and Analytics Options for  
Business Users

By David Stodder



Sponsored by



AUGUST 2011

# TDWI CHECKLIST REPORT

## Data Discovery and Unified Information Access: Expanding BI and Analytics Options for Business Users

By David Stodder



1201 Monster Road SW, Suite 250  
Renton, WA 98057

**T** 425.277.9126  
**F** 425.687.2842  
**E** info@tdwi.org

tdwi.org

### TABLE OF CONTENTS

- 2 **FOREWORD**
- 2 **NUMBER ONE**  
Understand requirements for self-directed data access and analysis.
- 3 **NUMBER TWO**  
Put users on a fast path to information insight.
- 3 **NUMBER THREE**  
Give users the power to perform deep analysis.
- 4 **NUMBER FOUR**  
Give users rich data experiences through visualization.
- 4 **NUMBER FIVE**  
Expand data reach into spatial data and unstructured content.
- 5 **NUMBER SIX**  
Improve data discovery through unified information access (UIA).
- 5 **NUMBER SEVEN**  
Give employees better information for improving processes.
- 6 **NUMBER EIGHT**  
Define IT's role in governing and provisioning data.
- 6 **NUMBER NINE**  
Develop and apply best practices for UIA and data discovery.
- 7 **ABOUT OUR SPONSORS**
- 8 **ABOUT THE AUTHOR**
- 8 **ABOUT TDWI RESEARCH**

© 2011 by TDWI (The Data Warehousing Institute™), a division of 1105 Media, Inc. All rights reserved. Reproductions in whole or part are prohibited except by written permission. E-mail requests or feedback to info@tdwi.org. Product and company names mentioned herein may be trademarks and/or registered trademarks of their respective companies.

### FOREWORD

Demand for business intelligence (BI) and analytics continues to be strong, but expectations are changing. At one time, business users who knew of BI understood and accepted that it involved a steep learning curve, batch queries on limited historical data, static reports, and long waits for IT to build new applications. Now that BI is topping research surveys as the most desired application, emerging expectations are for easy-to-use, self-service, and graphical environments that let users work interactively with and explore different types of data. Users want rapid application development and deployment capabilities for meeting changing requirements. From their dashboards and other kinds of portals, they want to access and analyze not only structured data, but also unstructured content.

Organizations struggle to meet such elevated expectations with most existing BI systems. This has created a prime opportunity for data discovery tools, which offer a fresh approach to data access, reporting, and analytics. Equally interesting are unified information access (UIA) tools, which share many of the qualities of the data discovery category but focus on integrating BI, search, and analytics for both structured and unstructured data. Leading tools in both categories give users better control over their environments, including the ability to test analytics in sandboxes and do rapid proof-of-concept systems before wider deployment. Many use new technologies for in-memory analytics and extended or hybrid SQL for different types of data such as spatial or geographical location data.

This Checklist Report focuses on helping organizations understand how data discovery and UIA tools address new requirements for data access, reporting, and analytics, particularly for nontechnical users in lines of business and operations. The Checklist examines how these new tools can help organizations add breadth, depth, speed, and flexibility to users' pursuit of information insight from both structured data and unstructured content. The report closes with guidance for developing best practices and ensuring that IT data governance and provisioning concerns are addressed while meeting expectations for greater user self-service and flexibility.



### NUMBER ONE

#### UNDERSTAND REQUIREMENTS FOR SELF-DIRECTED DATA ACCESS AND ANALYSIS.

Organizations have traditionally had a tough time extending business intelligence and data analysis capabilities to nontechnical users working in operations and lines of business (LOBs). The standard approach, in which IT drives BI development and deployment, can be fraught with difficulty. When IT seeks to gather user requirements, developers try to ascertain what users need in terms of data, reports, and analysis capabilities. However, this process can overlook two central issues: users' need to explore information and their role in authoring reports and assembling dashboards.

Users are not uniform. If enterprise BI tools lack flexibility, they leave users who are at different levels of maturity unsatisfied. Even individual users can have varying degrees of experience or clarity about what kind of reports, analysis, and visualization they need for different data requirements. The old adage that you don't know what your second question will be until you have the answer to the first one has never been more true. For some objectives, parameterized reports set up by IT are fine; for others, users may want to be in full control of the design, building, and sharing of analyses and best practices. The need for flexibility is a big reason why there's strong interest in tools that enable self-service BI and analytics, which can allow users to shape BI to their roles in business processes rather than depend on IT.

Across the spectrum, users increasingly need—and expect—an integrated view of structured data and unstructured content. To complement the “facts” delivered by structured data that tell them what's happening in their areas of concern, unstructured content from sources such as customer comments, social media/sentiment, news, forms, and more can help users see context and understand why things are happening the way they are. Yet, once again, it's difficult for IT developers to customize unstructured content analysis for each user; it may be better to give users tools to do it for themselves.

With data discovery technologies maturing, IT and LOBs should change requirements-gathering procedures to promote opportunities for users to perform self-directed data access and analysis.



### NUMBER TWO

PUT USERS ON A FAST PATH TO INFORMATION INSIGHT.

In most industries today, the fastest belong to the spoils. Poor information flow prevents organizations from being first to market with new products and services. If marketing functions have to wait several months for the development of BI applications that allow them to analyze data about campaign performance, opportunities can be lost. Supply chains are suboptimal when managers lack access to data; product flows are out of sync with demand; and customer feedback has little impact. Managers in operations cannot allocate precious labor resources to respond to immediate and projected needs.

Increasing speed to insight should therefore be a primary objective of deploying data discovery tools. In LOBs and operations, users need information to make daily decisions. If BI is too tightly controlled by IT and tools are geared to power users and experienced analysts, the rest of the user population has to function with incomplete and less timely information. Unstructured content comes to them haphazardly, preventing users from gaining a complete picture. Speed is not just about real-time data updates; it is also about giving users a faster path to all the information they need.

The development of BI dashboards has been a huge boon, bringing data access, display, and analysis to a wider population. BI and data discovery tool vendors are competing to provide users with the ability to personalize how they look at metrics, key performance indicators (KPIs), charts, and more. However, with expectations for dashboards rising, IT developers are under pressure to keep up with the pace of business for access to larger and more varied data volumes. Self-service capabilities that enable users to tailor dashboards can be crucial to reducing development backlogs.

Three key issues in evaluating whether dashboards are increasing speed to insight include relevance, completeness, and depth. Dashboards should give users information, both structured and unstructured, that is relevant to carrying out their responsibilities. Dashboards should bring together access and visualization of internal and external data sources, including unstructured content, spatial data, and documents, so that users gain a complete view. Finally, dashboards should enable users to access detailed data and take their discovery deeper.



### NUMBER THREE

GIVE USERS THE POWER TO PERFORM DEEP ANALYSIS.

“Discovery” is about finding out what you do not know. Thus, a key aspect of data discovery is going beyond limited views to analyze detailed data and unearth significant and unexpected insights. Strategic objectives that require business agility, predictive knowledge of customer behavior, or deep understanding of financial performance make it imperative for LOB and operational users to engage in more powerful data analysis. While some aspects of data and text mining will remain the special preserve of Ph.D. “data scientists,” data discovery tools are bringing these capabilities to nontechnical users.

BI users have traditionally investigated data through ad hoc queries. Dashboards have made it easier for users to engage in complex analytics by enabling them to use graphical means rather than actually writing SQL. Data warehouses support ad hoc querying, but IT organizations manage these queries carefully to guard against requests that might overload systems or attempt access to sensitive data. In addition, since existing ETL procedures and data designs are unable to account for every possible query, additional development may slow the path to insight.

Some BI and data discovery tools are improving on standard methods by automating or encapsulating ETL, data quality, and related steps that are necessary to serve up data for queries. These tools help shield users from data complexity and reduce hand-coding by IT developers. Other tools exploit large memory to move detailed data closer to users, enabling them to test predictive models and perform iterative analysis in memory and on their own.

In-memory analytics are powerful, but there is also a level of hype. Organizations need to examine how in-memory analytics works for the types of queries and workloads their users generate. If data volumes are too large or data must be updated frequently, other options that enable access to live data may be better. The best of all worlds is to have both options at your disposal, because you need both. Use in-memory options where appropriate and allow live data access when users need it.

Analytics is all about discovery, but there’s no one-size-fits-all solution. Look carefully at user requirements to match workloads with appropriate technology.

**✓ NUMBER FOUR**

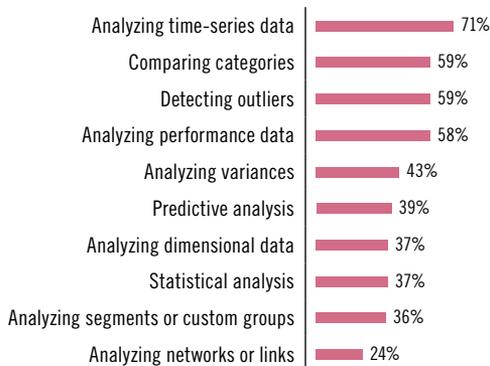
**GIVE USERS RICH DATA EXPERIENCES THROUGH VISUALIZATION.**

Data visualization is much more than just eye candy; it is critical to user productivity. Classic BI tools were initially slow to expand the menu of possible ways in which users could present and analyze data in their dashboards. This gave the new class of data discovery tools an opening. Although mainstream BI systems are improving, one of the defining features of data discovery tools remains the ease with which users can create data visualizations for reporting and analysis and try new ones without involving IT developers.

Many of these tools take advantage of large memory, compression techniques, and 64-bit, multi-core processors to ensure support for the intensive processing activity generated by interactive visual reporting and analysis. Leading data discovery and BI tools also employ the latest development software on the front end, including Java applets, ActiveX controls, Dynamic HTML, and rich Internet application environments to enable sophisticated visualizations and let users personalize how they blend data with other content.

In TDWI's 2011 *Visual Reporting and Analysis Best Practices Report*, two-thirds of organizations surveyed said that data visualization improves user productivity to a high degree. Nontechnical users are able to see patterns and trends that would either take too long or require greater data analysis expertise to find. Thus, visualization plays a key role in enabling these users to analyze big data sets and apply best practices for creating repeatable, effective, and actionable displays through the use of business rules and visualization logic. Figure 1 shows the various types of analyses for which users perceive data visualization as most important.

**Rate the degree of usefulness of visualization for the following analytical tasks.**



*Figure 1. Based on 210 respondents who ranked visualization's importance as "high."*

**✓ NUMBER FIVE**

**EXPAND DATA REACH INTO SPATIAL DATA AND UNSTRUCTURED CONTENT.**

Business opportunity and data variety often go hand in hand. The more types of data sources that become available, the more users can expand the scope of reporting and analysis beyond the limits of their data silos. Marketing, finance, fraud detection, risk management, and other functions can get closer to a 360-degree view of customers, patients, products, fraudulent activity, and more. The numbers alone cannot provide this view. Because people and process interaction involves e-mail, call logs, documents, social media, maps, and multimedia, tapping unstructured content sources is vital.

Spatial or geographical data is one type that is growing in importance as organizations look for new insights. "Micro" marketing, which involves a sharp focus on demographic data and location information, is becoming critical to campaigns across channels, including the Web. Retail decisions involving real estate or supply chains need spatial data and the ability to integrate it with other sources to optimize store locations and make decisions about what to display and where in each store. Data about spatial positioning is important to managing logistics, equipment, energy resources, agricultural sites, and more. Mobile devices are generating location data that could be valuable for spatial analysis.

BI and data discovery tools that enable analysis of spatial data can give nontechnical users in LOBs a physical dimension to go along with time series and other types of analysis for daily decisions and process roles. Given that spatial data is often consumed through looking at maps and three-dimensional representations, data visualization capabilities are important. In choosing tools for spatial analysis, organizations should examine whether the products use proprietary methods or implement spatial versions or extensions of SQL and common industry data access standards. In addition, organizations should look at how the data is indexed, loaded, and checked for quality to ensure that when users analyze and share spatial information they know to what degree it is complete and trusted.

In keeping with the goal of self-service for nontechnical users, spatial analysis capabilities should be easy to use. The capabilities should also integrate well with users' other BI and analysis tools.

### NUMBER SIX

#### IMPROVE DATA DISCOVERY THROUGH UNIFIED INFORMATION ACCESS (UIA).

Users lose productivity and miss important trends and relationships when they have to spend time moving between different tools to access structured data and unstructured content. Shuffling through documents is particularly problematic, since it has often been left to users to organize, access, and share them. Add in the explosion of online content together with voice-of-the-customer records and more, and it's easy to see why the unstructured world presents both problems and opportunities for reporting, visualization, and analysis.

BI systems grew up in a technology environment that historically focused exclusively on structured data, leaving the realm of unstructured content to search engines and content management systems. In recent years, BI tools have begun to integrate these worlds by providing search functionality in the suites. Data discovery tools have pushed further to make search and text analysis more integral to the user experience by binding these functions together with the ability to process SQL statements. Industry analysts have named the category of tools that integrate search, text analysis, and BI “unified information access (UIA).”

Bringing previously separate information worlds together helps reduce the chaos of multiple tool interfaces and adds power to analysis. Users can employ search and text analysis to reach out to sources that may be highly relevant but are not represented in the BI system's metadata. Using inverted indexes, UIA can find information faster than BI systems when users are in discovery mode. Users looking at KPIs, for example, may reach outside to consult Web sites that provide context for structured data. Tools that offer more sophisticated concept classification and faceted search enable users to go beyond the limits of available taxonomies and tailor content discovery to higher-level concepts.

Organizations should evaluate how well tools integrate structured and unstructured worlds. Search, text analysis, and BI functionality may be available through a single interface, but more advanced systems let users search BI sources and use SQL-based query languages to get results from content repositories. With each step toward true UIA functionality, organizations can reduce the time and difficulty involved in working with different types of information.

### NUMBER SEVEN

#### GIVE EMPLOYEES BETTER INFORMATION FOR IMPROVING PROCESSES.

Poor information flow to employees carrying out responsibilities in business, legal, manufacturing, or supply chain processes is the Achilles' heel of many organizations. Processes are almost always both data- and content-intensive, often to the point that employees have to spend too much time looking for information. Without tools that provide integrated views and analysis of structured and unstructured information, they cannot react quickly enough to find the root causes of process delays. Managers held accountable for processes struggle to meet objectives for creating more agility, reducing costs, and increasing efficiency.

BI, data discovery, and UIA tools can play a critical role in giving managers and frontline employees information visibility according to their role-based permissions and requirements. The “discovery” aspect is important because, although under pressure to solve a problem, employees don't always know what they are looking for. They may need to search through unstructured content and perform exploratory analysis on near-real-time data. To track processes and make adjustments, nontechnical employees need easily understandable dashboards that can alert them when changes occur or when exceptions to normal or benchmarked activities are detected. Visualization options also prove valuable for identifying patterns and seeing the significance of data changes.

A longstanding knock on BI systems is that they have not been well integrated into process management. This has been due to institutional and funding issues as much as technology impediments, which generally come down to BI's inability to function at the speed of processes. However, technology is improving, and more organizations are recognizing the value of BI and analytics for continuous process improvement. BI, data discovery, and UIA systems are evolving to provide complete environments for analysis, reporting, and measurement as well as tools for designing workflows that can provide documentation for process execution. These workflows can then be integrated with process workflows created with other tools.

With economic pressures showing no signs of abating, organizations need to seize opportunities to reduce waste and run more efficient and effective processes. Organizations should evaluate the potential of BI, data discovery, and UIA tools for process improvement.

### NUMBER EIGHT

#### DEFINE IT'S ROLE IN GOVERNING AND PROVISIONING DATA.

Data discovery requires more than a little independence from IT. In large part, this is not about politics; it's about reality. IT resource limitations create backlogs for BI and analytics applications. IT must manage data for everyone, not just individual users; it must manage data assets as a whole to ensure security, availability, reliability, performance, and adherence to regulations. "Shadow" data stores, spreadsheets, and LOB application data silos are a continuing problem for IT management when data in those sources must be shared, refreshed, and governed as part of the organization's assets.

LOB and operational users, however, need freedom and flexibility to blend or mash up data from different sources on the fly without IT involvement. They do not want to wait for the next data warehouse to be built, or to access only sources in existing data schemas. Users also need to access unstructured content from both inside and outside the organization and cannot be limited to existing taxonomies. Newer, less standard data sources, including spatial and geographical location data and Hadoop files, beckon for more complete and advanced analysis.

Organizations need to avoid repeating past errors that have led to data chaos and tensions between business users and IT. Self-service data discovery and UIA tool implementations should be a benefit to organizations, not a problem. The solution is to bring IT and LOB leaders together in BI centers of excellence or competency centers to iron out data governance policies and work together on improving and updating information architectures.

Data and content owners, focused on their own concerns, often pay too little attention to the drawbacks of incomplete or out-of-date metadata and taxonomies. Yet these resources are vital to the rapid development of applications and services. IT can provide leadership on these issues as well as policy matters regarding how data is used. With severe consequences hovering over many organizations if they expose data to privacy or security breaches, IT has a responsibility to govern data assets. Governance, however, need not stand in the way of users realizing the benefits of self-service UIA and data discovery.

### NUMBER NINE

#### DEVELOP AND APPLY BEST PRACTICES FOR UIA AND DATA DISCOVERY.

Users are hungry for new ways to analyze and share information. Data discovery and UIA tools can help users leverage growing volumes of data and content inside and outside their organizations. They can set tools up with less IT involvement and gain from flexible and interactive interfaces to delve deeper into the data. Leading tools enable users to break down walls between structured and unstructured data to open up opportunities for fresh insight.

Organizations should evaluate tools that take advantage of technologies such as in-memory analytics, spatial data analysis, and advanced search functionality. Here are five objectives for best practices aimed at giving users speedier paths to complete insight:

- **Make it easier and faster for users to tap new data sources.** Users derive value when they can perform discovery analysis with new sources of data and content. Evaluate tools and systems that bring new sources online for rapid analysis of specialty and time-sensitive sources, such as files containing results for Web site tests.
- **Let users personalize discovery and visualization.** The popularity of smart phones and tablets demonstrates that users want personal choice in how they consume information. Evaluate tools that can help users personalize their interaction with data and content.
- **Choose tools that leverage the value of big data.** In social media, on the Web, and in large transaction systems and data warehouses, enormous volumes of potentially valuable data and unstructured content flow into and around organizations. Data discovery and UIA tools take advantage of technology changes to enable nontechnical users to realize value more rapidly. Evaluate how well tools scale for big data analysis.
- **Establish managed self-service.** User freedom and flexibility are vital, but IT has a critical role in provisioning and governing data. Create centers of excellence to enable business users and IT to work out data governance and collaborate on metadata and taxonomies.
- **Find and document repeatable user analysis patterns.** LOBs and operations can avoid reinventing the wheel by developing information workflows for shared reporting, visualization, and analysis, which can then be integrated with business processes.

**ABOUT OUR SPONSORS**



Solving business problems for nearly a quarter-million users worldwide, Alteryx is driving the global revolution in agile business intelligence with analytics by empowering business leaders with insight. Through smart, extensible solutions from the desktop to the cloud, the Alteryx software platform delivers the fastest, most comprehensive consumer, business, and market insights to the *Fortune* 500, mid-market companies, government, and academia. Alteryx desktop, enterprise server, and cloud solutions integrate spatial intelligence into enterprise workflows, seamlessly scaled across local, regional, and global markets. Inspiring ingenuity since 1997, Alteryx, Inc., is headquartered in Irvine, CA, with its Technology Center in Boulder, CO. For more information, please visit [www.alteryx.com](http://www.alteryx.com).



Critical decisions and processes depend on information that is created and stored across the organization in many formats—databases, spreadsheets, SharePoint, documents, e-mails, call logs, and more. Attivio solves this long-standing problem of information isolation by integrating and connecting these diverse sources to help our customers not only understand *what* is happening, but also provide critical context so they can understand *why* it is happening.

Attivio's dynamic approach to unified information access (UIA) integrates search, business intelligence, data warehousing, and analytic capabilities to deliver insight that matters. Our Active Intelligence Engine™ (AIE™) supports SQL and simple, search-style queries to retrieve information via reports, dashboards, and custom interfaces. AIE features rapid development, incremental scalability, information-driven workflow, text analytics, real-time updates and permissioning, and a small footprint that can be easily embedded in other products and applications.

Business and IT leaders, system integrators, and independent software vendors partner with Attivio to address today's global business imperatives—uncovering insights that drive innovation and reveal new opportunities, mitigating risk and controlling cost. For more information, please visit [www.attivio.com](http://www.attivio.com).



Tableau Software, a privately held company in Seattle, WA, builds software that delivers fast analytics and rapid-fire business intelligence to everyday businesspeople. Our mission is simple: help people see and understand data. Tableau's award-winning products integrate data exploration and visualization to make analytics fast, easy, and fun. They include Tableau Desktop, Tableau Server, Tableau Digital, and the free Tableau Public.

Tableau understands the needs of businesspeople, nontechnical and technical alike, when it comes to retrieving and analyzing large volumes of data. As a result, Tableau has already attracted over 65,000 licensed users in companies from one-person businesses to the world's largest organizations. For more information, please visit [www.tableausoftware.com](http://www.tableausoftware.com).

### ABOUT THE TDWI CHECKLIST REPORT SERIES

TDWI Checklist Reports provide an overview of success factors for a specific project in business intelligence, data warehousing, or a related data management discipline. Companies may use this overview to get organized before beginning a project or to identify goals and areas of improvement for current projects.

### ABOUT THE AUTHOR

**David Stodder** is TDWI Research Director for Business Intelligence. As an analyst, writer, and researcher, he has provided thought leadership on key topics in BI, analytics, IT, and information management for over two decades. Previously, he headed up his own independent firm and served as vice president and research director with Ventana Research. He was the founding chief editor of *Intelligent Enterprise*, a major publication and media site dedicated to the BI and data warehousing community, and served as editorial director there for nine years. With TDWI Research, Stodder focuses on providing research-based insight and best practices for organizations implementing BI, analytics, performance management, and related technologies and methods. You can reach him at [dstodder@tdwi.org](mailto:dstodder@tdwi.org).

### ABOUT TDWI RESEARCH

TDWI Research provides research and advice for business intelligence and data warehousing professionals worldwide. TDWI Research focuses exclusively on BI/DW issues and teams up with industry thought leaders and practitioners to deliver both broad and deep understanding of the business and technical challenges surrounding the deployment and use of business intelligence and data warehousing solutions. TDWI Research offers in-depth research reports, commentary, inquiry services, and topical conferences as well as strategic planning services to user and vendor organizations.