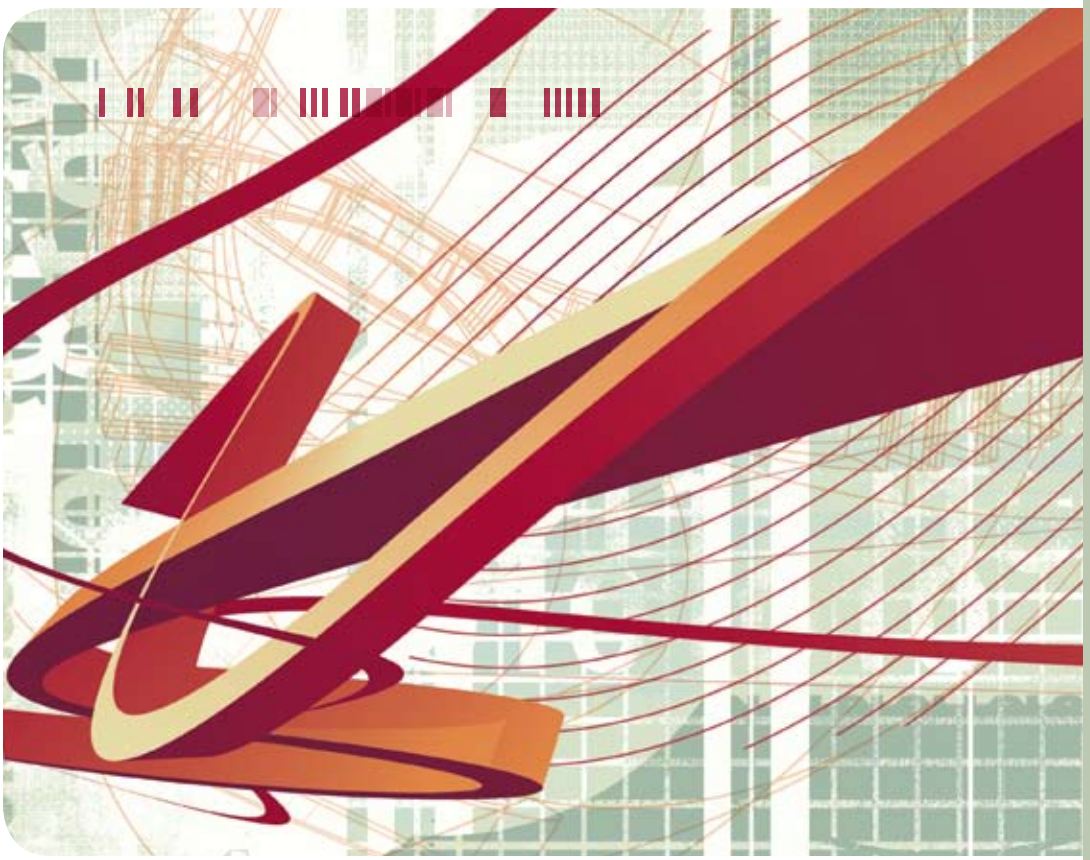
SECOND QUARTER 2008
TDWI BEST PRACTICES REPORT

DATA GOVERNANCE STRATEGIES

Helping your Organization Comply, Transform, and Integrate

By Philip Russom



www.tdwi.org

tdwi
THE DATA WAREHOUSING INSTITUTE

Research Sponsors

Business Objects, an SAP company

Collaborative Consulting

DataFlux

Exeros

Informatica Corporation

SAP

SAS

Trillium Software

DATA GOVERNANCE STRATEGIES

By Philip Russom

Helping your Organization Comply, Transform, and Integrate

Table of Contents

About the Author

PHILIP RUSSOM is the senior manager of TDWI Research at The Data Warehousing Institute (TDWI), where he oversees many of TDWI's research-oriented publications, services, and events. Before joining TDWI in 2005, Russom was an industry analyst covering BI at Forrester Research, Giga Information Group, and Hurwitz Group. He also ran his own business as an independent industry analyst and BI consultant, and was contributing editor with *Intelligent Enterprise* and *DM Review* magazines. Before that, Russom worked in technical and marketing positions for various database vendors. You can reach him at prussom@tdwi.org.

About TDWI

The Data Warehousing Institute, a division of 1105 Media, Inc., is the premier provider of in-depth, high-quality education and training in the business intelligence and data warehousing industry. TDWI is dedicated to educating business and information technology professionals about the strategies, techniques, and tools required to successfully design, build, and maintain data warehouses. It also fosters the advancement of data warehousing research and contributes to knowledge transfer and the professional development of its Members. TDWI sponsors and promotes a worldwide Membership program, quarterly educational conferences, regional educational seminars, onsite courses, solution provider partnerships, awards programs for best practices and leadership, resourceful publications, an in-depth research program, and a comprehensive Web site (www.tdwi.org).

About TDWI Best Practices Reports

This series is designed to educate technical and business professionals about new business intelligence technologies, concepts, or approaches that address a significant problem or issue. Research for the reports is conducted via interviews with industry experts and leading-edge user companies, and is supplemented by surveys of business intelligence professionals.

To support the program, TDWI seeks vendors that collectively wish to evangelize a new approach to solving business intelligence problems or an emerging technology discipline. By banding together, sponsors can validate a new market niche and educate organizations about alternative solutions to critical business intelligence issues. Please contact TDWI Research Director Wayne Eckerson (weckerson@tdwi.org) to suggest a topic that meets these requirements.

Acknowledgments

TDWI would like to thank many people who contributed to this report. First, we appreciate the many users who responded to our survey, especially those who responded to our requests for phone interviews. Second, our report sponsors, who diligently reviewed outlines, survey questions, and report drafts. Finally, we would like to recognize TDWI's production team: Jennifer Agee, Bill Grimmer, Denelle Hanlon, Deirdre Hoffman, and Susan Stoddard.

Sponsors

The research for this report was sponsored by Business Objects, an SAP company; Collaborative Consulting; DataFlux; Exeros; Informatica Corporation; SAP; SAS; and Trillium Software.

Research Methodology and Demographics

Report Scope. This report is designed for business and technical executives who are responsible for planning and implementing a program for data governance (DG). This report helps organizations worldwide successfully navigate the unknown waters of DG by presenting the best practice techniques and common pitfalls involved in starting and sustaining a DG program. It identifies and evaluates common starting points and strategies, with an emphasis on the cross-functional nature of DG.

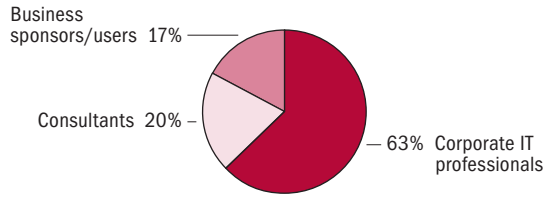
Survey Methodology. Most of the market statistics presented in this report are based on the report's survey. In November 2007, TDWI sent an invitation via e-mail to the data management professionals in its database, asking them to complete an Internet-based survey. The invitation also appeared on several Web sites and newsletters, and 424 people completed all of the survey's questions. From these, we excluded the respondents who identified themselves as academics or vendor employees, leaving the completed surveys of 394 respondents as the primary data sample for this report. Additional data comes from a Technology Survey run at a TDWI World Conference in August 2007.

TDWI also conducted telephone interviews with numerous technical users and their business sponsors, and received product briefings from vendors that offer products and services related to the best practices under discussion.

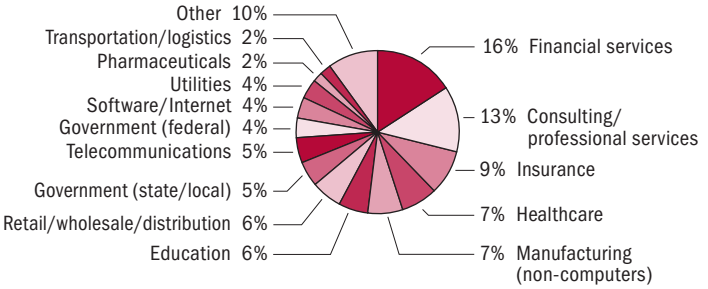
Survey Demographics. The wide majority of survey respondents are corporate IT professionals (63%), whereas the remainder consists of consultants (20%) or business sponsors/users (17%). Because DG is often led by executives, this survey drew a lot of respondents with high-level job titles such as director, vice president, and chief officer. But note that most of these executives identified themselves as IT professionals, meaning that they manage IT. We asked consultants to fill out the survey with a recent client in mind.

The financial services (16%) and consulting industries (13%) dominate the respondent population, followed by insurance (9%), healthcare (7%), manufacturing (7%), education (6%), retail (6%), and miscellaneous industries. Most respondents reside in the U.S. (66%) or Europe (14%). Respondents are fairly evenly distributed across all sizes of companies and other organizations.

Position

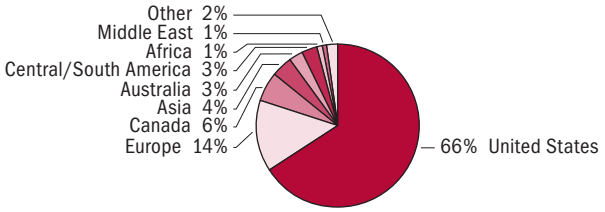


Industry

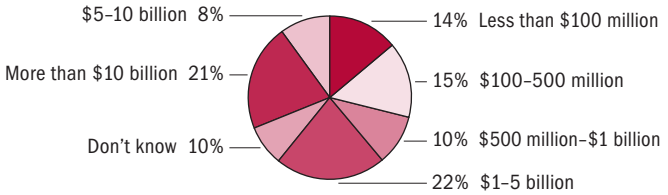


(Above, the "other" category consists of multiple industries, each represented by less than 2% of respondents.)

Geography



Company Size by Revenue



Based on 394 survey respondents.

Executive Summary

Complying with regulations and extending data integration are common initiatives that gain from data governance.

Anytime data crosses an organizational boundary, it should be governed, whether you're sharing data among business units internally or publishing data to customers, partners, auditors, and regulatory bodies externally. Furthermore, we now live in the "age of accountability," which (among other things) demands stricter oversight for data usage, quality, privacy, and security. Organizations are under renewed pressure to ensure that compliance and accountability requirements are met as the scope of data integration broadens. In response to this situation, many organizations are turning to data governance, which establishes policies and procedures for sharing data, as well as improving data's quality, structure, and auditability.

Data governance enables technology change and business transformation.

Furthermore, a goal of some data governance programs is to enable an organization to treat data as an organizational asset. Achieving this goal demands many interim goals, most involving dramatic change. For example, data governance transforms an organization's data, its data management technology, who owns the data, and how the organization uses data. Sweeping changes and business transformations like these need a central organizational structure such as a data governance committee or board, staffed with both business and technology people. The board must institute and enforce policies and procedures for data management and business use of data. And data governance is best coordinated with IT governance and corporate governance.

Data governance intersects with most data-driven business initiatives and technical implementations.

Once under way, data governance affects data-driven business initiatives like compliance, business intelligence, customer relationship management, and business transformations (such as reorganizations and mergers and acquisitions). When executed broadly, data governance becomes a part of almost all data management practices, including data quality, integration, warehousing, administration, architecture, and lifecycle management. Organizations typically choose a starting point from among these initiatives and implementations, then incorporate others later based on pain points and priorities.

Software automation for data governance today comes mostly from selected functions in certain tools.

The execution of data governance is all about the four Ps: People collaborate to create procedures and policies, and all that comes together into a data governance process. In other words, most data governance tasks are purely interpersonal and organizational. Yet, software automation is important, because it can potentially give data governance greater speed, accuracy, and scalability. Applications dedicated solely to data governance are rare. Software automation for data governance is already available through selected functions in tools for data quality, data integration, metadata management, and master data management. As users better define their requirements for data governance, software vendors will no doubt supply new functions and tools.

Data governance is new and complex, so confusion abounds.

Given the complexity of data governance and its many influences, it's no surprise that confusion abounds. Although many data management and business professionals have experience with data governance, few of these professionals have practiced it deeply. There are many approaches to data governance, as seen in the diverse best practices of user organizations and the array of products and services offered by vendors.

This report from TDWI Research clears the confusion by drilling into the business initiatives, technical implementations, and cross-functional organizational structures with which data governance intersects. It also quantifies the state of data governance adoption and describes some of the technologies and vendor products that can help automate data governance. All this information is tailored to assist business and technical managers in planning and implementing a sustainable data governance program.

Overview of Data Governance

Definitions of Data Governance

Data governance is hard to define because it's still new and evolving. Each organization tailors data governance to its needs and abilities, and DG is practiced both in isolated pockets as well as on an enterprise scale. Furthermore, DG is inherently a cross-functional program that involves a mix of technology and business people—plus their IT systems and business processes—and the mix varies greatly.

Even so, here's a definition that covers almost all the components and goals of data governance:

> Data governance (DG) is usually manifested as an executive-level data governance board, committee, or other organizational structure that creates and enforces policies and procedures for the business use and technical management of data across the entire organization. Common goals of data governance are to improve data's quality; remediate its inconsistencies; share it broadly; leverage its aggregate for competitive advantage; manage change relative to data usage; and comply with internal and external regulations and standards for data usage. In a nutshell, data governance is an organizational structure that oversees the broad use and usability of data as an enterprise asset.

That's a mouthful. So, here's a rule of thumb that's easy to remember:

> DG usually boils down to some form of control for data and its usage.

The catch is that "control" has multiple meanings that are somewhat at odds:

- **DG may tighten control to limit data access.** This is true when data governance is driven mostly by compliance goals, especially data security and privacy.
- **DG may ease control to expand data integration.** Most DG boards provide procedures through which a team can request access to data owned by another team. Ironically, this eases the control of data to assist initiatives that rely on broadly integrated data, like business intelligence (BI) and customer relationship management (CRM).
- **DG may define controls that improve the content of data or dictate its structure.** For example, data flows through many IT systems and departments, so improving the quality of data (whether physical or semantic) is a cross-departmental affair that DG can manage. Likewise, enterprise data architecture seeks to tweak the structure of multiple databases for the sake of easier database management or data integration. DG can define standards that dictate consistency for data structures and data definitions.
- **The level of control can vary.** For example, strict governance is typical of federally mandated compliance, whereas loose guidance is typical of data architecture standards. Or, a multi-divisional corporation may demand strict governance for data at the headquarters level so data yields a unified view of total corporate performance, yet merely provide loose guidance for individual implementations so local organizations can satisfy local requirements.

DG is about finding the right definition and level of control.

"In our consulting practice, we have participated in data governance initiatives that evolved from either grass-roots data management or executive fiat," said David Loshin, president of consultancy Knowledge Integrity, Inc. "In one situation, the need for standardizing shared data representations drove the 'bottom-up' development of a governance infrastructure, leading to a federated data standards governance framework. In another situation, the introduction of a consolidated enterprise

EXPERT COMMENT

DG programs may progress bottom-up or top-down.

application suite was expected to be accompanied by data governance as directed 'top-down' by senior management. Whether bottom-up or top-down, both cases posed common challenges in communication, standardization of concepts, and establishing operational processes for governance. On the technology side, data quality, metadata, and policy management were success factors."

Critical Attributes of Data Governance

These definitions help us understand the goals and actions of most data governance programs. To fill out the rest of the picture, here are other attributes of successful programs. Note that all are core assumptions of this report, and all are discussed in detail later:

People, policies, and procedures constitute the DG process.

- **Data governance is mostly about the "four Ps."** These are seen most clearly in the DG board, where people work together to establish and enforce policies (or rules) defining which data is subject to governance, as well as the allowable access and usage of such data. Procedures provide a structure for reviewing and acting on requests for data access, data improvement, and other changes. *People*, *policies*, and *procedures* all combine to enable a larger DG *process* (see Figure 1). The four Ps explicitly remind us that DG is mostly about people collaborating to establish a DG process that accommodates the needs of all their business units (and external entities, when appropriate), but with priority to enterprise goals.

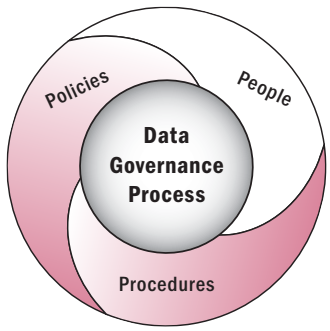


Figure 1. The data governance process consists of people, procedures, and policies.

Don't do DG in a vacuum.

- **Data governance must coordinate with other forms of governance.** Don't forget that data governance is but one form of governance. TDWI Research has interviewed people who've made DG work in isolation (say, just for BI or data quality). But, in the long run, DG should coordinate with other forms, especially IT governance and corporate governance.
- **Data governance doesn't govern data directly.** The term "data governance" leads us to believe that we are governing data directly. But the truth is that we're governing how data is accessed and used via business initiatives, as well as defined and managed via data management infrastructure. This explains why DG is increasingly a component of these initiatives and infrastructures.

DG intersects with business initiatives and technology practices.

- **Data governance intersects with business initiatives.** An assumption of this report is that data governance touches many different business initiatives, especially those that are data-driven, like compliance, BI, CRM, and business transformations. DG is often a subset of these initiatives, and is increasingly a critical success factor for them.
- **Data governance intersects with data management practices.** When executed broadly, DG influences almost all data management practices, including data quality, integration, warehousing, standards, administration, architecture, and lifecycle management. DG typically

requires that adjustments be made in these practices, in support of the policies developed by the DG board. While tools dedicated to DG are rare today, some data management tools can automate some actions of DG.

- **A successful DG program strikes a pragmatic balance among competing goals.** For example, there's a prominent need for balance between compliance goals that limit data access and business integration goals that expand data access. Other opposing goals include business versus technology, data content versus data usage, strict governance versus loose guidance, and departmental versus enterprise data ownership. Most DG programs start in one carefully bounded area that serves a single goal (like DG just for BI, compliance, or master data management [MDM]), so the balancing act is not immediately apparent. Striking an appropriate balance becomes a critical success factor as the program expands to govern more data sets, data usage scenarios, and data management practices. Such balances are difficult to attain and maintain without the executive mandate, central policy making, change management procedures, and cross-functional collaboration of data governance.

DG is a balancing act.

Why Data Governance Now?

There are many reasons why organizations should initiate or expand DG programs now:

- **The current "age of accountability" demands compliance.** And punishments for non-compliance are severe, ranging from customer flight and revenue loss to fines and jail terms. Firms are under unprecedented pressure to control data usage according to internal policies for data security and privacy, as well as external regulations like Basel II, HIPAA, and SOX. Assuring compliance is an early-phase goal of most DG programs.[1]

Compliance is a pressing problem, and DG is a prominent part of the solution.

- **Compliance and business intelligence demand high-quality, auditable data.** Organizations need to improve the quality of data that goes into public documents, especially regulatory reports. Furthermore, report auditability—i.e., recording the lineage of report data—is crucial to surviving an audit, regardless of who the auditors are. And one of the most common questions asked by report consumers internally is: "Where did this data come from?" Today, the quality of report data is a high priority for most DG programs, whereas auditability is a lesser priority.
- **Improving data quality is a cross-functional imperative.** Since a DG board is cross-functional by nature, it's an ideal organizational structure to effect improvements that span multiple business units. Although data quality focuses mostly on physical data, master data and metadata need improvement, too. This is why many data quality and master data management initiatives are supported by a cross-functional DG board.

Solving data quality problems on an enterprise scale demands DG.

- **Data integration (DI) implementations cast an ever-widening net.** This is true whether DI is analytic (feeding a data warehouse), operational (consolidating database instances) or cross-business (sharing data with partners). DG can both limit these implementations to assure compliance and liberate them to reach more data sources and targets. DG can also assist by providing data exchange standards and procedures for data access and improvement requests.
- **Data governance reduces the risk incurred during business transformations.** DG is imperative in firms that experience regular transformations such as reorganizations, mergers and acquisitions, and initiatives that involve data as an enterprise asset (typically linked to CRM or sometimes BI). These transformations require extensive changes in data ownership and data structure. DG can manage the changes while assuring compliance.

DG improves practices and reduces risk for business integration and transformation.

[1] Visit TDWI's White Paper Library (www.tdwi.org/WP) to download the IT Audit Checklist Series, which goes in depth into several issues mentioned in this report, including information security, data privacy, IT governance, and change management.

USER STORY

Though not common, DG can focus on data security.

"We started our data governance program a few years ago by focusing on HIPAA compliance and other data privacy issues," said the director of data governance at a prominent US retailer. "But then the data breaches at TJX become public knowledge, and that was a turning point. To be 100% sure we couldn't get hit like that, we refocused data governance on information security." In March 2007, The TJX Companies disclosed that 45.6 million credit and debit card numbers were siphoned from its servers over an 18-month period by unknown hackers. "We can't afford to get 'TJXed,' and now we feel confident that we won't."

Benefits and Barriers

In the Internet-based survey for this report, TDWI Research asked survey respondents to identify data attributes and data-driven business actions that would benefit from data governance. Their responses rank the likelihood of these potential benefits, as shown in Figure 2:

Data quality is the primary beneficiary of DG.

- **Data quality is the leading perceived benefit of DG.** Sixty percent of survey respondents selected the quality of data as a DG benefit. This is no surprise, since most DG programs include a beefy data quality component. Related quality areas ranked lower in survey respondents' perceptions, namely the quality of master data (23%) and the quality of metadata (20%), suggesting that users are more concerned about the quality of physical data than semantic data. Even so, second place went to consistent data definitions (58%); this involves semantic data, as well as ensuring that physical data is consistent with documented definitions.[2]
- **Data integration is also a prominent beneficiary of DG.** This is demonstrated in Figure 2, where data as an enterprise asset (52%) and decision making based on data (38%) bubbled up near the top. Other data integration issues ranked in the middle, namely the sharing of data (22%) and visibility into the enterprise via data (20%).
- **Compliance issues vary in terms of perceived benefits.** For example, appropriate use of data (33%) and accountability for data use (25%) rose into the top third of the rankings. Other compliance issues ranked in the bottom third, namely data security (16%), data in regulatory reports (12%), data lineage (12%), and data privacy (10%).
- **The cross-functional nature of DG is also a benefit, though not a top priority.** These issues ranked in the middle of Figure 2, including collaboration among teams (26%), change management processes for data use (18%), and business transformations (17%).

Despite potential benefits, survey respondents also perceive several barriers to data governance:

DG requires change, which people find threatening.

- **Turf wars and intransigence are the leading barriers to DG.** More than any other barrier, 68% of survey respondents pointed to data ownership and other territorial issues as the major stumbling blocks (see Figure 3). Related issues—like resistance to change or transformation (42%) and resistance to accountability (31%)—ranked as mid-level problems.

Education about DG can remove barriers.

- **Users need to study data governance before tackling it.** The second leading barrier in the survey is a lack of understanding of governance (61%). Related issues that could be solved through education include non-sustainable executive sponsorship (36%) and a lack of business justification (29%).
- **Cross-functional DG is a benefit, but only if you have cross-functional experience.** This shows up in Figure 3 as two prominent barriers: a lack of cross-business unit coordination (53%) and inexperience with cross-functional initiatives (28%).

[2] The intrinsic relationship between data quality and data governance is explained in the TDWI Best Practices Report *Taking Data Quality to the Enterprise through Data Governance*, available online at www.tdwi.org/research/reportseries.

- **Technology can be a barrier to DG.** Several respondents pointed to a poor state of data management infrastructure (42%) as a possible problem. In the interviews TDWI conducted for this report, users reported that certain DG goals—typically for data quality and integration—couldn't be met until implementations were deployed or revised.

Which of the following would benefit from data governance? (Select five or fewer.)

Figure 2. Based on 1,960 responses from 394 respondents.

What would be barriers to data governance in your organization? (Select five or fewer.)

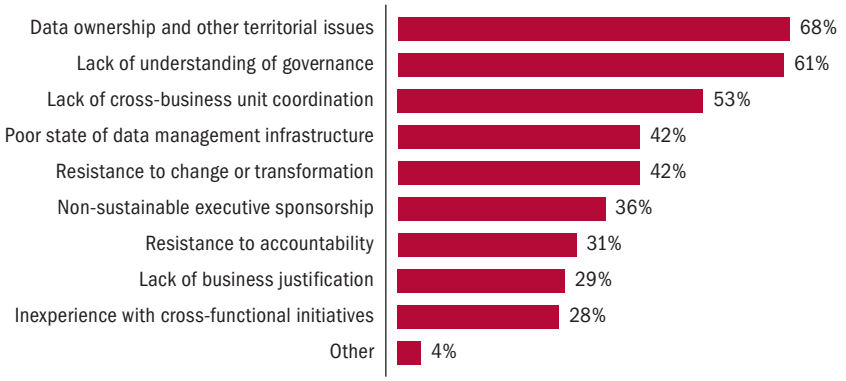


Figure 3. Based on 1,547 responses from 394 respondents.

"We have over 100 IT systems managing HR data, and we share HR data with over 150 partnering firms for payroll, outsourcing, staffing, healthcare, life insurance, and various benefits," said the systems consultant for HR data governance at a U.S. retailer. "Our main strategy for data governance in this context is to achieve privacy by limiting the amount of data that's shared internally and

USER STORY

Data privacy is about limiting access and complying with HIPAA.

externally. We're always looking for ways to tighten up application and data access internally, plus consolidate HR data into fewer, centralized locations. And we make sure that the data files generated for our partners include only the data that an individual partner needs, plus encryption. We also investigate our partners to ensure that they guard the data. And all that has to comply with HIPAA. Our HR data governance program coordinates closely with our financial data governance and corporate governance programs."

The State of Data Governance Programs

In an effort to quantify the state of DG programs, TDWI Research asked survey respondents about their programs and their personal involvement with them:

DG is new and rare, but poised to proliferate.

- **Data governance is still very new.** That explains why deployed programs for data governance are still rare. Only 15% of survey respondents reported reaching deployment (see Figure 4). Though deployed, it's a safe bet that most of these programs are still in early stages.
- **More DG deployments are coming.** One-third of surveyed organizations are already in a design or implementation phase (33%), with many more considering a DG program (40%).
- **DG is definitely on corporate radar screens.** A whopping 88% of respondents in Figure 4 have made some form of commitment to DG, although most commitments are in an early stage. Few organizations have no plans for data governance (12%).

DG experience is still moderate, so success levels are, too.

- **DG experience is common, despite DG's newness.** Most of the respondents for the question in Figure 5 report that they participate in (42%) or manage (27%) a DG program. Even so, few organizations have gone deep into DG yet, so experience cannot be deep.
- **DG success is mediocre, so far.** In Figure 6, almost half of respondents claim a medium success rate, with one-quarter or less claiming high or low. Few respondents report a very high or very low rate. The mediocre success rate for DG isn't bad—or surprising—given that its best practices are not yet fully known and disseminated.

What's the status of your organization's DG program?

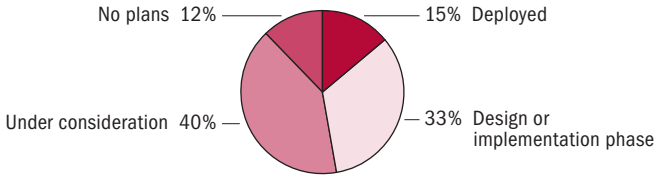


Figure 4. Based on 394 respondents.

Do you currently manage or participate in a data governance initiative?

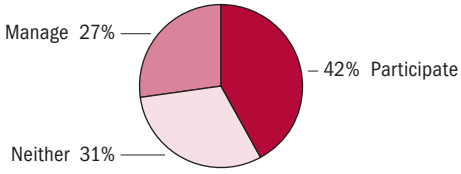


Figure 5. Based on 394 respondents.

Rate the success of your organization's data governance initiative.

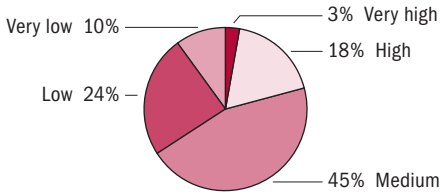


Figure 6. Based on 273 respondents who have DG experience.[3]

The Role of DG in Initiatives and Implementations

Data governance intersects with many different business initiatives and IT implementations. The intersection, of course, results from a focus on data that DG shares with data-driven business initiatives like compliance, BI, CRM, business transformations, global spend analysis, accurate reporting and forecasting, and so on. Likewise, policies and procedures established by a DG program influence almost all data management implementations, including those for data quality, data integration, data warehousing, database administration, enterprise data architecture, and MDM.

The possible intersections between DG and various initiatives and implementations strongly determine the attributes of a DG program:

The scope of DG is determined by the initiatives and implementations governed.

- **Organizations choose a starting point from the intersections.** For example, most DG boards start by governing a compliance initiative or a data quality implementation, because these are common pain points that require immediate attention.
- **The number of intersections addressed define scope.** In other words, your list of currently governed initiatives and implementations defines the current scope of your DG program.
- **The intersections addressed affect DG details.** For example, you might staff the DG board based on whose data (or whose data usage) is currently being governed, as well as who is best equipped to define and enforce policies for the current list.

Knowledge of initiatives and implementations governed at other organizations can help you plan your DG program.

- **DG details affect initiatives and implementations governed.** As requests for data access or improvement are reviewed by the DG board, approved requests lead to a longer list of systems, data sets, and business units involved. And a DG board may proactively expand into an area when doing so is crucial to enforcing its policies and data standards (as with mergers and acquisitions).
- **Knowing the possible intersections helps you predict where DG will be required or useful.** Regardless of where you start, some initiatives (like compliance) or implementations (especially data quality) will inevitably fall under the scope of your DG program.
- **One way to plan DG program phases is to prioritize the list of possible intersections.** When possible, a DG board should plan a reasonable sequence of future initiatives or implementations for DG, based on each one's level of pain, return on investment, and the willingness of involved parties to participate.

[3] Due to branching in the online survey, only respondents with DG experience (based on their response to the question in Figure 5) were allowed to answer this question.

Data Governance for Data-Driven Business Initiatives

A Technology Survey run at a 2007 TDWI World Conference asked attendees: "What types of business initiatives do you think should be guided by data governance?" From their responses, a priority order arises:

Users feel that BI, compliance, and transformation initiatives need DG, in that order.

- **BI is the leading candidate for DG, at least with BI professionals.** Predictably, the survey respondents—who are mostly BI professionals—selected business intelligence far more often than other business initiatives as a prime target for data governance (88% of survey respondents in Figure 7). This makes perfect sense, because many business users consume corporate data that's delivered to them via reports and other media generated by BI systems. When DG raises the quality and consistency of data, decisions based on the data likewise improve, whether tactical or strategic. When BI data guides sensitive decisions—like whose mortgage gets foreclosed or which customers receive credit or which suppliers receive preferred status—BI data and its usage demands governance to avoid discrimination and to assure compliance with corporate and external regulations.
- **Compliance issues are pressing, even among BI professionals.** Just about every business unit and technology team is currently being affected by compliance (66%) and related issues like data privacy (69%). Compliance is inherently linked to regulations for data usage, so the link to data governance is inevitable.
- **Business transformations transform data, too, so DG is required.** A business transformation involves dramatic changes to the structure of a corporation, including how its IT systems are aligned with and owned by organizational units. The most obvious examples of business transformations (47%) are mergers and acquisitions (34%) and reorganizations (28% in Figure 7). However, some types of sales and marketing campaigns (26%) are transformational when they force cross-business-unit data integration or data quality implementations. Likewise, initiatives that redefine data as "an enterprise asset" force similar transformations relative to data's content, structure, and ownership.

  The point is that business transformations are inherently risky because of the extensive amount of change involved. The change management procedures of DG and other forms of governance help reduce risk and keep the transformation focused on agreed goals. Most DG boards are chaired by an executive whose mandate helps the transformation clear the hurdles of organizational resistance.

USER STORY:

DG can enable business transformations.

"UMB Bank was built on a strong customer-focused strategy. However, over time with new products coming to market, the customer experience become product centric and not relationship centric," said Kevin Kramer, senior VP of enterprise sales at UMB. "Since we knew we had to provide our officers the complete picture of the customer's experience with UMB, it became imperative that we streamline access to our customer data and institute a fundamental change of our sales process." To achieve this transformation, UMB broke down some organizational silos that segregated customers and provided a mechanism for aggregating customer information. UMB succeeded thanks to a data governance program. "I wanted IT to be seen as a pure fulfillment organization, not as the sponsor of data governance," said Kanon Cozad, senior VP of application development at UMB. Today, the data governance program is directly sponsored by the highest level of management, which focuses on business change, with technology providing a supporting role. UMB's data governance program continues to foster positive change, but now also drives business growth. The deep and positive impact of UMB Bank's business transformation won them the 2007 TDWI Best Practices Award for data governance.

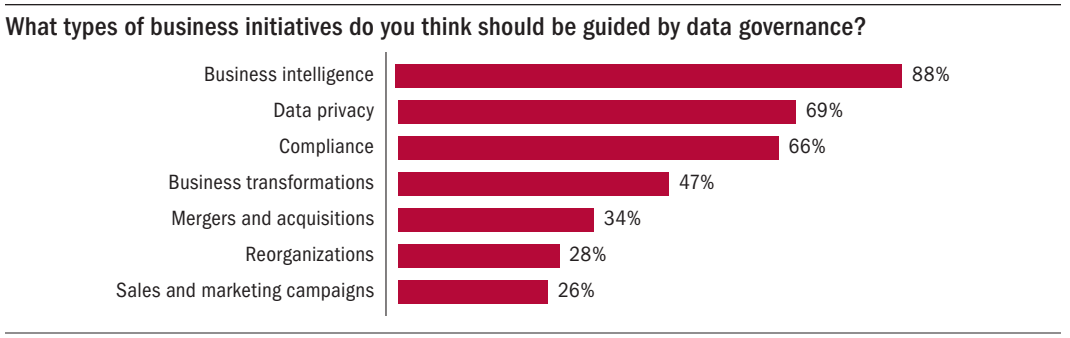


Figure 7. Based on a TDWI Technology Survey of August 2007, 119 respondents.

Data Governance for Data Management Implementations

The TDWI Technology Survey mentioned earlier also asked: "Which data management practices do you think should be guided by data governance?"

Note that the question suggests that a DG board would govern how data management professionals design implementations. Based on interviews with users conducted for this report, DG boards govern data-driven implementations to a limited degree, but only to ensure that implementations support DG policies and other business requirements. Think of this as guidance on the specification level, not the design level. With that in mind, it's possible that respondents misread the question as: "Which data management practices do you think can best support data governance policies?" After all, the DG process is usually a two-way street: DG influences data management practices, and they in turn support DG policies. Regardless of which direction survey respondents had in mind, their responses reveal a clear priority order:

Users feel that data integration and quality implementations need DG, for both physical and semantic data.

- **Data integration and quality go hand-in-hand with DG.** At the top of their picks, survey respondents strongly linked data integration and data quality to data governance (83% and 82%, respectively, in Figure 8). Data integration implementations are ripe for DG controls, because they "use" data by accessing and transporting it, often across organizational boundaries. Data quality has become almost synonymous with data governance because it's a critical success factor in DG-driven initiatives for compliance, business transformation, and business integration. Furthermore, the natural synergy between data integration and data quality practices has in recent years led practitioners to mix the two freely in seamless implementations. Now, a new synergy mixes them with data governance.
- **Master data and metadata need DG, too.** Most discussions of data governance focus on controlling and improving physical data, as above where data integration and quality are identified as top priorities. Yet, survey respondents also strongly linked DG with semantic data in the forms of master data management (MDM) and metadata management (75% and 74%, respectively, in Figure 8). In particular, MDM benefits from DG (and vice versa) because a DG program provides a collaborative organizational structure for developing the consensus-driven data definitions that MDM is all about. Furthermore, the mandate of the board ensures that data definitions will be applied once defined. In fact, TDWI Research has interviewed multiple users who started a DG program explicitly to support an MDM implementation.[4]
- **DG can affect data models and architecture.** Though not a high priority and still evolving today, survey respondents recognized this issue by selecting data warehousing and enterprise data architecture (both 56% in Figure 8). To a lesser degree, DG can also affect operational

[4] For more information about data governance for MDM, see the TDWI Best Practices Report *Master Data Management: Consensus-Driven Data Definitions for Cross-Application Consistency*, available online at www.tdwi.org/research/reportseries.

database administration (26%). As an example, remember that data warehouse teams have, for decades, collaborated with business people to assure that the data model of the warehouse represents the business accurately and in ways that support reporting and analysis requirements. This process resembles DG, and some of the users interviewed talked about how it became a model—or even the actual basis—for their current DG program.

From another angle, consider that DG policies sometimes dictate changes to data models, so they comply with data standards (whether physical or semantic) or are more conducive to data integration and quality operations. Here, we've crossed the line into modern concepts of enterprise data architecture. The Internet survey for this report revealed that people with the title of data architect are often chairs or co-chairs of DG boards, which shows that the goals of data architecture are progressively mixing with DG—and even leading it.

USER STORY:

Master data management is a common goal of data governance.

"About 10 years ago, I designed our first global SAP system," said a systems architect at a leading cell phone manufacturer. "We had a nice template, and everything was cool at first. But the template, as rolled out in other systems, allowed for master data variations across instances. After a few years of rollouts, we started seeing problems from inconsistent data definitions. In particular, the finance department really needed more consistency for faster book closing. So, a finance manager set up a central master data group, from scratch, which grew to 15 people in the first year, though focused on governing only one system. Over the last five years, the group has extended to govern many other systems, involving 60–70 people. We started with around 19 categories of master data—mostly financials and customers—and we recently extended to HR data. Product data is next. Data governance was an easy sell for us, because of the consistencies in global operations and speed of financial closings gained from master data management."

Which data management practices do you think should be guided by data governance?

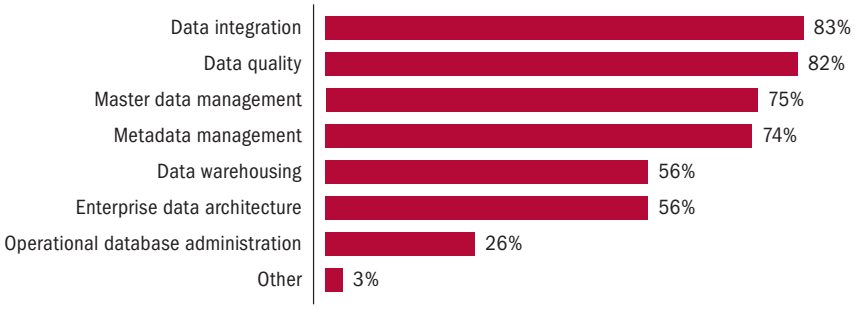


Figure 8. Based on a TDWI Technology Survey of August 2007, 117 respondents.

Prioritizing the Phases of Data Governance

Early phases of most DG programs focus on data quality, integration, and compliance.

The online survey for this report asked users with DG experience to identify business initiatives and data management implementations that were addressed (or will be addressed) in the early, later, and future phases of their DG programs (see Figure 9). Their responses reveal organizational priorities and indicate where successful DG programs start and where they go over time. Readers of this report should use this information to plan the phases of their programs.

- **Most users begin their DG programs by focusing on data quality and integration.** Data quality is the unqualified leading priority for the early phases of users' DG programs, followed closely by initiatives and implementations that involve data integration, like business intelligence, data migration/consolidation, and making data an enterprise asset.

- **Compliance is another common starting point.** This includes related issues like data security and data privacy.
- **Later phases of DG touch just about everything.** For every initiative or implementation listed in Figure 9, approximately 30% of survey respondents selected it. This even distribution of responses suggests that middle phases reach a maturity where DG touches most relevant initiatives and implementations, whereas early and future phases have selective priorities.
- **A few issues are postponed for future phases of a DG program.** These include master data management (MDM), application consolidation, and mergers and acquisitions. Although these are not an early priority for the majority of users surveyed, TDWI Research interviewed a minority of users who founded their DG programs to address these issues, especially MDM.

USER STORY

Governing data as it is entered prevents problems later and enables new uses of data.

"One of my responsibilities is what I call operational data governance," said David Woods, the director of global data management in Johnson & Johnson's consumer products division. "We have to assure data quality, and we know that most data problems originate when users of transactional applications enter data or when a data integration routine automatically loads data for them. Operational data governance includes multiple levels of data validation that correct, enhance, and sometimes convert data to our standards as it enters a system, so we need not fix it downstream later.

"Good validation requires sophisticated rules, and that's where governance comes in," Woods continued. "Our governance process allows us to develop consensus-based data standards and invoke changes to applications and databases so as to introduce better validation and standardization. In order to be competitive in today's environment, the firm must be able to react quickly and seamlessly to organizational acquisitions, and our data governance policies help us quickly repurpose data from acquired companies into our standards. And the firm depends on reporting and analysis of operational data on a global basis, sometimes in real time, which our clean and standardized operational data enables."

Among the following business initiatives and data management practices, which were involved with your data governance program in its earliest phases versus more recent phases? Which will be involved in future phases?

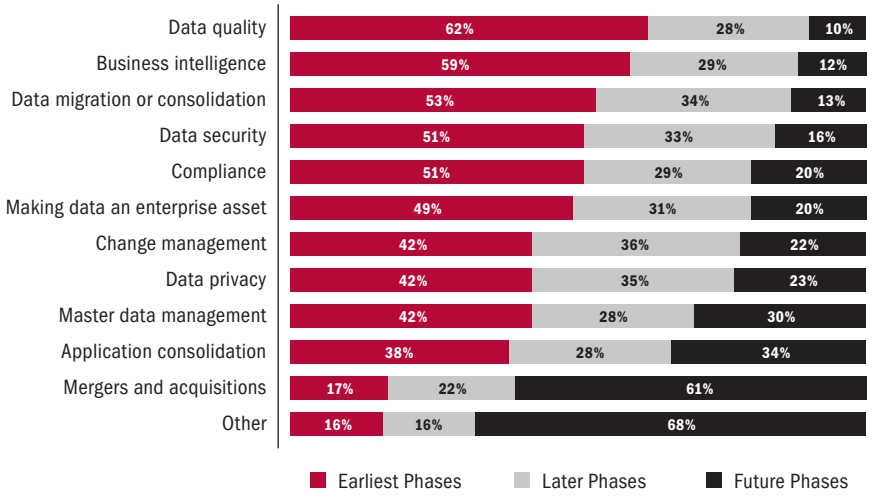


Figure 9. Based on 273 respondents who have DG experience.[5]

[5] Due to branching in the online survey, only respondents with DG experience (based on their response to the question in Figure 5) were allowed to answer this question.

The Three Pillars of DG

The many business initiatives and data management implementations mentioned above constitute a daunting list that's tough to remember. Let's give the list some structure and make it more memorable by boiling it down to the three main pillars of an enterprisewide program, namely: compliance, transformation, and integration. (See Figure 10.)

Most DG goals boil down to compliance, transformation, and integration.

These three pillars are a generalization that represents the most common goals for data governance programs. Each can be a starting point for a new DG program, and all three relate and share dependencies in a mature, enterprisewide program. Let's drill into the three pillars.

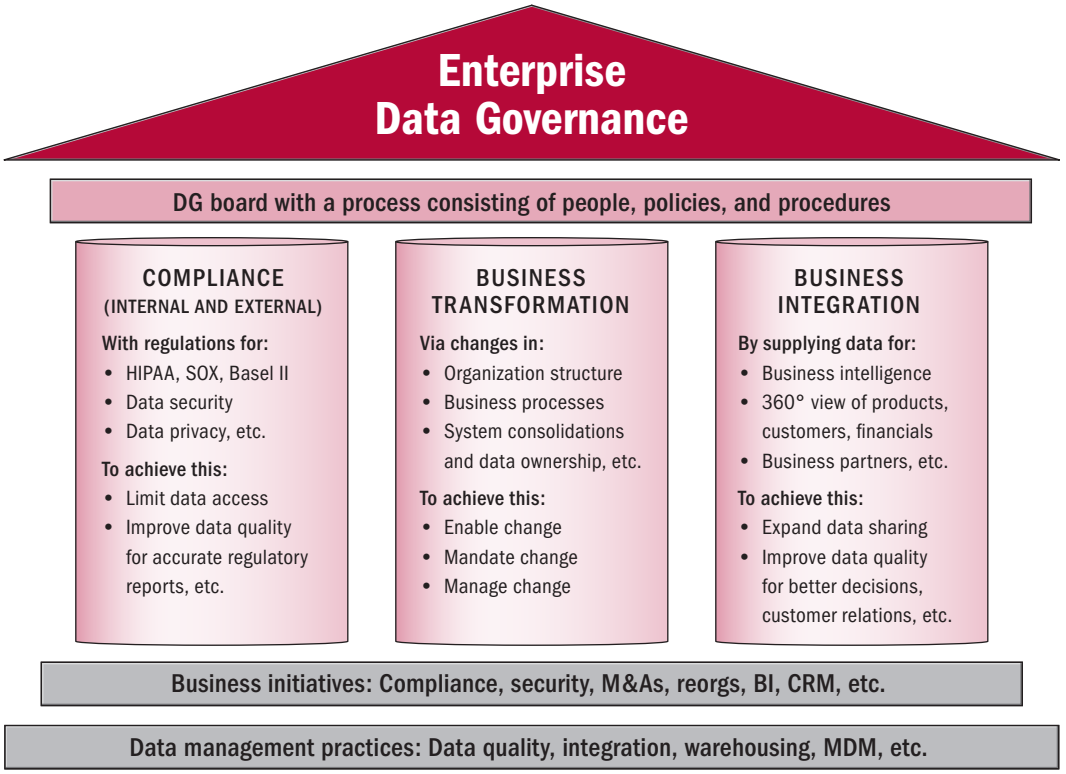


Figure 10. Compliance, transformation, and integration are the three main pillars of data governance.

Compliance

At one end of the spectrum, many organizations initiate data governance programs because of pressing compliance issues that impact data usage. These issues are diverse, involving internal policies (for data security and privacy), legislated regulations (Basel II, HIPAA, and SOX), and standards for data exchange (EDI, HL7, SWIFT, etc.). An organization may begin with a subset of data-related compliance—typically HIPAA and other data privacy regulations—then expand to SOX compliance, regulatory reports, and data security. Note that achieving compliance relative to data often involves *limiting* the number of people and applications that may access certain data, as well as how they may use the data.

A data governance board assists by interpreting what each of these compliance issues means to the organization. The interpretation should be expressed as policies and procedures for data usage that must be followed by specific members of the organization. When compliance requires regulatory

reports, governing the quality of report data takes priority. Sometimes DG mandates that an audit trail for report data be kept in case the organization is audited. You can see that compliance, data governance, and data quality are strongly related, which is why all three are part of a single program in many organizations.

Business Integration

At the other end of the spectrum, some organizations begin by governing data that's shared broadly through a variety of data integration and application integration technologies. Technology aside, the point is to achieve business integration through data integration. While shared data is subject to compliance, the focus here is on *expanding* data access and integration. The goal of business integration is to enable data-driven business activities like BI, CRM, and cross-business data exchange (whether between business units of the same firm or partners in a supply chain). Sharing data broadly gives these initiatives faster and better decisions, customer relations, supply chain operations, and so on.

The data governance board can help expand data integration by providing procedures through which one organization can request access to another's data. Since downstream data practices like data warehousing and customer data integration suffer from the poor quality of upstream data, DG procedures should enable these teams to request improvements in upstream data sources such as operational applications. A DG board can establish data standards for exchange. And some firms go so far as to establish a DG-controlled integration competency center, which provides an infrastructure for data integration.

Business Transformation

Business transformation forces dramatic changes in data usage.

In the middle, transformation is a goal unto itself, as well as an enabler for the goals of compliance and integration. The meaning and degree of transformation varies, but it always involves some kind of change. For instance, to comply with the data security requirements of SOX, most U.S. corporations have this decade changed who can access which data. Other firms have changed the ownership of data, as they move toward using "data as an enterprise asset." Most improvements in data quality require that operational application owners and users change their systems and how they're used. And we've all seen numerous business transformations—especially reorganizations and mergers and acquisitions—that force dramatic changes in data ownership and usage.

Change is a constant with DG, and change must be managed.

Regardless of the degree of change seen in these examples, all benefit from data governance as a change management mechanism. Users whom TDWI Research interviewed for this report regularly pointed out that their data management and business process changes would not have happened without a data governance board to approve the changes and then enforce them across all business units. Several interviewees went so far as to create a DG board first, before attempting business or technical initiatives that require considerable changes to data ownership, quality, and usage.

Again, DG-controlled business transformation is a goal unto itself (especially with reorganizations and mergers), as well as an enabler (in the form of change management) for the data governance goals of compliance and integration. That's why, in Figure 10, enterprise data governance both arches over the three pillars and reaches into the middle one. Furthermore, the figure positions transformation in the middle, because a DG board can serve as a broker between compliance's need to limit data access and integration's need to expand it. In other words, DG can help expand data sharing without compromising data compliance goals.

Organizational Issues for Data Governance

The "four Ps" mentioned early in this report tell us that the DG *process* is mostly about *people* collaborating to establish *policies* for data usage and *procedures* for proposing changes to data usage, management, and ownership. The four Ps are a handy mnemonic for the basic components of DG, yet each P itself has components (as described in the following section of this report). Pulling all these components together into a cohesive process generally demands an organizational structure—a DG board or equivalent—that's related to other structures and staffed appropriately.

Cross-Functional Collaboration as a Requirement for Data Governance

Data governance is cross-functional by nature in that its process spans multiple types of people, business units, business initiatives, and data management practices. Hence, a DG program and its organizational structure tend to be cross-functional in multiple ways:

DG is cross-functional in multiple ways.

- **Across IT and business.** In a lot of ways, governance is a "killer app" for IT-to-business alignment. The term itself brings together IT (data) and business (governance). In some firms, the DG program is a continuation of IT-to-business alignment, which DG can reinvigorate. In others, DG has other priorities, but is designed to leverage prior work in alignment. With that in mind, the staff of a DG board needs to be a mix of IT and business people.
- **Across business units.** A goal of many DG programs is business integration, which shares data aggressively across multiple business units, whether these are within a single enterprise or strewn across multiple enterprises. Similarly, DG may guide business transformations, which often involve a change in business unit structure. The policies and procedures of DG should apply to all business units and their data, and the DG board may be staffed to represent many of the units.
- **Across business initiatives and data management implementations.** As explained earlier, DG is often tied directly to specific business initiatives and data management implementations to ensure that they fulfill business and compliance requirements. When a single DG program is tied to all these, it becomes a collaborative hub for sharing best practices involved in supporting business goals through data management practices.

EXPERT COMMENT

DG facilitates the interaction and negotiation that's fundamental to cross-functional decision making.

"Good data governance requires so-called decision-making bodies," said Jill Dyché, a partner at Baseline Consulting. "A single, standalone council can be insufficient. There will likely be multiple teams, departments, or groups who will want a say in the decisions around data. There is a significant amount of interaction and negotiation involved, and incumbent committees should be engaged. The most successful data governance efforts are those that are systemic—facilitating the interaction between multiple decision-making bodies, but ultimately ensuring that the data governance process is 'baked in' to new business and IT initiatives.

"Data governance can be complex and culturally specific, so it must be explicitly designed," Dyché continued. "It cannot simply be mandated by an executive or an organization. There's no substitute for the iterative diverge-and-converge process of thinking it through, because that process fosters ownership and commitment, as well as driving clear tactics for moving forward in a tactical and sustained way."

Organizational Structures that Support Data Governance

Multiple organizational structures can support DG—not just a DG board or committee.

This report stresses the data governance board (or committee) as a common organizational structure for DG. But companies, government agencies, and educational institutions are fairly creative in finding or creating a home for DG. For DG to be effective, it needs an organizational structure that's

cross-functional and collaborative, and there are many of these around to provide precedence and a model for DG—and possibly a home. Anyone planning a DG program should learn from these, mimic their structures, and possibly borrow their resources.

- **BI and data warehousing teams.** Building a data warehouse with a BI solution atop it is inherently cross-functional, in that business requirements must be expressed in multiple components of the technical solution. Furthermore, the BI team itself is cross-functional, in that it commonly includes specialists in data integration, quality, and modeling, as well as reporting, analysis, and performance management.
- **Data quality and stewardship programs.** Early on, data quality specialists realized that improving data requires cooperation from many business units and IT teams. Hence, data stewards arose as hybrid personnel who foster collaboration among units and teams with the goal of finding and prioritizing quality improvements. Many firms begin with a stewardship program and expand it to cover general DG.
- **Enterprise data architecture groups.** Data architecture used to focus on modeling individual databases. The modern definition, however, is primarily about standards, validation, and lifecycle issues among large collections of databases, and secondarily about data integration, quality, and compliance issues specific to enterprise databases. Because of an intersection between the two, data architecture and DG may share staff.
- **Competency centers.** By definition, a competency center (or center of excellence, as it's sometimes called) is a team that provides centralized, shared services for technical implementations. Despite the technical focus, a competency center may also include business or hybrid personnel that foster cross-functional or collaborative roles. BI competency centers and data integration competency centers (both common among TDWI Members) are good examples of mixed staffing. A competency center can be a home for a DG board or coordinate closely with one. A few firms have gone so far as to create a DG competency center, which is mostly concerned with the process of DG, but can also handle technical implementations.

Data Governance's Relationship to Other Forms of Governance

Data governance is but one form of governance, and multiple forms of governance commonly coexist and coordinate with each other. For example, in a hierarchal relationship, corporate governance is a broad umbrella that sets enterprisewide policies; IT governance may be a subset under corporate governance, while DG is a subset of IT governance. Occasionally DG coexists and at peer level with BI governance or data stewardship. And DG may have "dotted line" responsibilities with other organizational structures, like a compliance board, the offices of the CIO or CFO, or any of the business initiatives, IT implementations, and cross-functional organizational structures mentioned earlier in this report.

Coordinate DG with other governance programs, but with minimal bureaucracy.

Some corporations (typically large, multinational ones) develop a complex hierarchy of governance boards. DG has a bureaucratic bent to it, and so it tends to flourish in corporate cultures that tolerate a certain level of bureaucracy. But a complex hierarchy is beyond the tolerance of most cultures, and the overhead of coordinating multiple boards can be a barrier to success.

The consensus among users is that DG must unquestionably be coordinated with IT governance and corporate governance. In fact, users surveyed think the coordination should be tight (50%) or moderate (46%). (See Figure 11.) A mere 4% think the relationship should be loose. The catch is to define a depth of coordination supported by appropriate communication mechanisms that will provide effective data governance without an undue bureaucratic burden.

How closely should data governance be coordinated with IT governance or corporate governance?

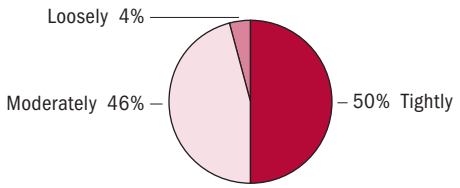


Figure 11. Based on a TDWI Technology Survey of August 2007, 116 respondents.

USER STORY

Flexible governance balances corporate control with project autonomy.

The 2006 TDWI Best Practices Award for data governance was awarded to IBM's internal CIO office for its flexible approach to governance. Flexible governance may sound like an oxymoron, but it's not. Flexibility and innovation are driving forces in IBM's business growth, and these are present in the enterprisewide governance model that IBM applies internally. The initial focus is on coupling strategic business goals with investments in information technology. Once investment plans are made, IBM's governance model provides consumable information management (IM) guidance, with flexibility for autonomous execution at the project level. This governance backdrop assures that the right guidance is available with a flexible level of control to the right players at the right time to ensure success.

Staffing the Data Governance Board

Populating the DG board with diverse people from diverse departments is critical to achieving the cross-functional collaboration required of data governance. To quantify DG board staffing, TDWI asked survey respondents with DG experience, "Who sits on your data governance board or committee?" (See Figure 12.)

Diversity is key to staffing a DG board.

- **IT personnel dominate the board.** This includes data warehousing or BI directors (56%), enterprise data architects (47%), and miscellaneous IT directors (25%).
- **Business managers.** Line-of-business (LOB) managers (56%) have a vested interest in the DG process, because they must communicate compliance policies to their direct reports. For example, when a DG board establishes policies for how end users should use a specific application, the LOB manager who owns the application or manages the end users should explain and police the policy. And LOB managers can accurately prioritize data quality and data sharing opportunities because they know which ones will yield a return on the investment.
- **Hybrid IT/business people.** Given the cross-functional and collaborative nature of the DG board's work, it's useful to have people with knowledge of both business and technology, like data stewards (48%) and business analysts (42%).
- **Upper management.** Despite their time-consuming jobs, CxOs sometimes sit on DG boards, as seen in survey responses for chief information officer (33%), chief finance officer (20%), and compliance officers (20%). The presence of high-placed executives (even if only emblematic) is key to establishing a mandate for the enforcement of DG policies.

Who sits on your data governance board or committee? (Select all that apply.)

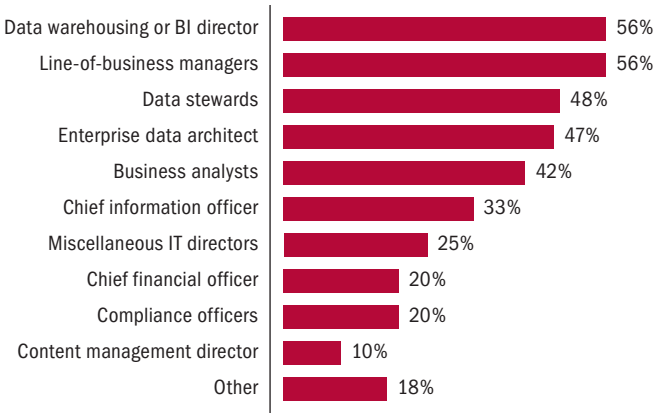


Figure 12. Based on 1,026 responses from 273 respondents who have DG experience.

Leading the Data Governance Board

It's clear that the staff of a DG board or similar organizational structure should include diverse people from diverse departments, so that the concerns of important or relevant departments are represented. But leadership for the board is a different matter, which begs the question: Who chairs or co-chairs the board? To quantify the issue, this report's online survey instructed respondents with DG experience to: "Enter the job title of the person (or persons) who lead your firm's DG board or committee." There were no prewritten answers to select; instead, each respondent typed an answer in his or her own words. A few of the 273 respondents entered multiple job titles, because their boards have co-chairs, resulting in 281 responses.

Directors dominate the DG board, especially those from IT.

- **The most common title for the DG board leader is director (27%).** And one-fifth of the directors had the phrase "data governance" in their job titles (6%), which is a sure sign that DG has arrived as an established corporate discipline. Other director-level job titles included the phrases "information technology," "business intelligence," and "data management," showing that technical disciplines dominate over business ones at the head of the DG conference table.
- **Manager job titles came in second (19%).** This includes diverse managers of both IT and lines of business. Related to these managers are the VPs (14%) who do the same jobs, but with a loftier title.

DG gets a strong mandate when led by a high-placed executive.

- **CxOs chair DG boards more often than anticipated (16%).** This is surprising given the tight schedules of chief officers and the fact that many delegate responsibility readily. Yet, given the technical side of DG, it's not surprising that most of these are CIOs or CTOs (9%), as compared to a few CFOs (4%) and CEOs (3%).
- **Data architects are common DG board chairs (9%).** A few survey respondents reported the DG board leader's job title as either "data architect" or "enterprise data architect," revealing that architects can be important leaders for the DG process.

Co-chairs can align business and IT in the DG process.

- **A few boards are led by two or more chairpersons.** Common co-chair combinations include a CIO with a CFO, a data architect with a BI manager, and a line-of-business manager (possibly at the VP level) with an IT manager or director. Note that co-chairs often represent IT and business, respectively, so the two remain aligned and balanced.

USER STORY:

The right business sponsor gives data governance a strong mandate.

"Our first attempt at data governance grew from a stewardship program firmly rooted in IT," said an enterprise data architect at a PC peripherals manufacturer. "It failed due to a lack of business participation, which was due to a weak business sponsor. In our company, the business owns the data, not IT, so the business must lead data governance. IT leading is like pushing a rope!" The turning point was a change in upper management. "We got a new CFO, who immediately set up a cross-functional governance council, including responsibilities for data governance and stewardship. Finally, we were equipped to address pressing issues in master data management and data mart consolidation. These are inherently cross-functional, and we didn't have an effective mandate until the new CFO's governance council came along."

Software Automation for Data Governance

The Role of Software Automation

As we've just seen, DG is inherently organizational and interpersonal. Even so, many of the tasks of the governance process—and many of the outcomes that result from enforcing a governance policy—can be automated (to some degree) with computer software and hardware. Common goals for any data-oriented software automation include the consistent, scalable, and auditable management, repurposing, and communication of information. With that in mind, there is noticeable overlap between the goals of DG and the capabilities of various data management tools. The overlap suggests that such tools, whether homegrown or vendor-built, can help automate DG processes and outcomes.

Over half of users think software can automate DG processes.

But is software automation for DG really possible? Over half of the respondents to this report's online survey said yes, a quarter said no, and the rest don't know, indicating that some kind of software automation for DG is possible (see Figure 13). TDWI's take on the situation is that some data governance tasks can be automated with software, and some can't, as explained below.

At the moment, software automation specifically designed for DG is somewhat light. A few selected functions or areas within certain types of data management and development tools support DG, but full-blown applications specifically for data governance are rare. TDWI suspects that the available automation for DG will increase over the next year or two, because lots of users are in the midst of defining their requirements for DG software automation, and vendors are already expanding data management products to address DG more directly.

Do you think data governance can be enabled by a homegrown or vendor-built software tool?

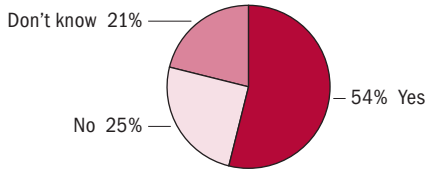


Figure 13. Based on 394 respondents.

Tool Types that Support Data Governance

In this report's online survey and user interviews, certain types of data management tools came to the foreground as both enablers for some data governance tasks, as well as software tools whose

implementations are affected by the policies of data governance. We can infer a number of things from users' input. First, some tools are highly conducive to data governance, and these are identified and discussed below. Second, among these, some tools are more conducive than others. This establishes a priority order (as represented in the order of the following discussion and in Figure 14) that can guide managers as they plan the phases of their DG programs. Third, the way that these preferred tool types are used reveals best practices and requirements for DG software automation. And all this can be represented in a basic technology stack for the software automation of data governance (see Figure 15). Furthermore, all these tool types are available as products from software vendors, and representative vendors and products are mentioned in the discussion below.[6]

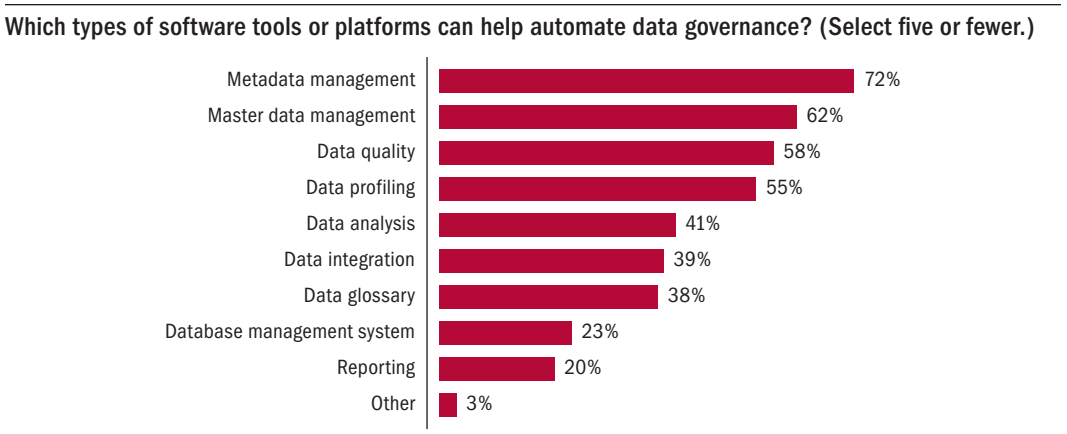


Figure 14. Based on 1,623 responses from 394 respondents.

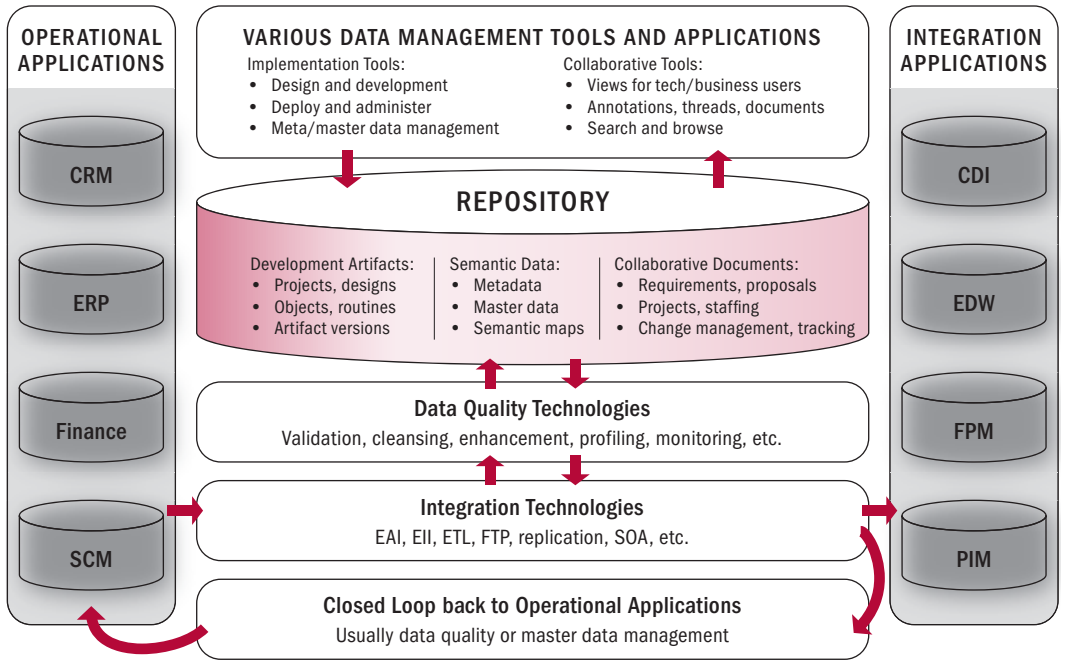


*Figure 15. A basic technology stack for the software automation of data governance.**

* Acronyms in Figure 15 stand for enterprise resource planning (ERP), customer relationship management (CRM), supply chain management (SCM), enterprise data warehouse (EDW), customer data integration (CDI), product information management (PIM), financial performance management (FPM), enterprise application integration (EAI), enterprise information integration (EII), extract, transform, and load (ETL), file transfer protocol (FTP), and service-oriented architecture (SOA).

[6] The vendors and products mentioned here are representative, and the list is not intended to be comprehensive.

Semantic tools

Management for semantic data is the leading requirement for DG software automation.

Users invariably point to the management of semantic data—namely metadata and master data—as a leading requirement for data governance. This involves the practices of metadata management (72%) and master data management (62%), which top the list of priorities in Figure 14. Semantic data is important because it helps build an inventory of governed data and define the meaning of governed data. The semantics may be expanded to define other attributes, like data's owners, sources, transformations, targets, quality state, dependencies, security issues, and so on. The data inventory and definitions are crucial to DG goals such as business integration, data quality, and auditability.

Metadata management and master data management functions are built into representative data integration tools (from Business Objects, an SAP company; Informatica Corporation; and SAS) and data quality tools (from the aforementioned vendors, plus DataFlux and Trillium Software). Exeros offers tools that discover cross-system business rules and metadata and identify data exceptions and anomalies that violate those business rules. Dedicated MDM applications are available from Exeros, SAP, and other vendors. While these aren't data governance tools per se, they make a significant contribution to the automation of data governance policies. As an interviewee put it, "NetWeaver MDM from SAP is an engine that helps you achieve data governance goals, although it's not a data governance tool."[7]

Central repository

Architecture for DG software automation is based around a central repository.

Managing semantic data is the top priority, yet an accoutrement of semantic infrastructure stands out: namely, the semantic data repository. Every metadata management and master data management solution has some kind of repository at its heart. This is true whether the tool is a metadata management tool or a master data management application. It's true whether the semantic solution stands alone or is embedded in a larger system, like an ERP application, data integration tool, or data quality tool. It's true whether the repository is a dedicated metadata management repository (typically object oriented) or a homegrown schema built atop a relational database management system. Regardless of its form, the repository is extraordinarily important, because it's the central, definitive source for semantic data and related data attributes. Figure 15 recognizes the repository's important role by placing it in the middle of the DG technology stack.

The central repository's importance has grown this decade as its breadth has increased. Even when a repository originated for a specific purpose like metadata management, users and vendors have stretched it to also manage other things, like master data, physical data, project documents, collaborative functions (e.g., versioning, Web views, discussion threads), and development artifacts (e.g., data flows, quality flows, hand-coded routines). Admittedly, the extended functions of repositories evolved to satisfy requirements for metadata-driven practices like data integration, quality, and profiling. But they lend themselves to the collaborative tasks of data governance, too.

Extended metadata repositories are built into data integration and quality tools from the vendors just mentioned. And, of course, many users implement their own repository, usually atop their corporate-standard brand of relational database management system. A single repository would be ideal, but multiple repositories are the norm, because repositories are built into most tools. A common best practice is to synchronize master and metadata across repositories, sometimes based on one of the repositories as a central system of record.

[7] For a complete survey of tools for MDM, see the TDWI Technology Market Report *Segmenting Master Data Management Solutions* (Q4 2006), available to TDWI Members at www.tdwi.org/research.

Data quality tools

The second leading requirement for DG software automation is data quality.

As we saw earlier in this report, improving the quality of data is a common goal of data governance, whether it's a goal unto itself or a supporter of other goals, like compliance, business integration, and business transformation. Given its ubiquity, you should consider data quality a firm requirement for the software automation of data governance. Users surveyed ranked data quality (58%)—and related functions like data profiling (55%) and data analysis (41%)—right after semantic data management issues (see Figure 14).

Data quality tools support a variety of quality operations for name-and-address cleansing, match-and-merge, deduplication, verification, enhancement, standardization, and so on. But they also include capabilities that have direct import to data governance. For example, data profiling helps a user discover data and quantify its state; this area within the tool originated for data quality purposes, but obviously has application in data governance. Likewise, data monitoring polls data after each run of a deployed data quality solution to assess whether the data is improving and to identify further opportunities for improvement; monitoring can assure compliance with DG policies, not just data quality standards. And the practice of data stewardship is so prominent in data quality implementations that most data quality tool vendors have created tools (or functions within a data quality tool) designed specifically for the steward as user. These tools accommodate the steward's level of technicality (which varies from very technical to mildly technical), helping the user discover data improvement opportunities, develop rules for data transformations, communicate these to technical developers, review reports based on data monitoring, and process exceptions with a mix of manual and automated methods. Again, all these capabilities have direct import to DG.

Representative data quality tools—which include related functions for profiling, monitoring, and stewardship—are available from Business Objects, DataFlux, Informatica, and Trillium Software.[8]

Cross-system data discovery and monitoring tools

A common limitation of data profiling and monitoring tools is that most operate on one database at a time, sometimes more narrowly on one table at a time. This is a problem in today's distributed data environments, where data dependencies link data elements across multiple tables and databases. When working with distributed data, users should look for tools that enable *cross-system* data relationship discovery and *cross-system* data monitoring, not just traditional data profiling and monitoring. Such tools should also include automation for deducing and documenting business rules and exceptions for the data relations and transformations that they find, since rules and exceptions are fundamental to assessing compliance with DG policies.

Representative tools for cross-system data relationship discovery and monitoring include Exeros Discovery and Exeros Validator.

Data integration tools

After semantic data management and data quality, data integration is the third most common tool type used to automate data governance (39% in Figure 14). Tools for data integration are an obvious requirement when software automation for data governance demands moving and transforming data, as it clearly does with diverse forms of business integration and BI. But data integration is less obvious when it supports a more primary DG requirement for compliance or business transformation. Of course, data integration itself should be governed so it is compliant with DG policies for data standards, security, and privacy.

[8] For a complete survey of data quality vendors and tools, see the TDWI Technology Market Report *Enterprise Data Quality Tools* (Q2 2006), available to TDWI Members at www.tdwi.org/research.

Data integration is a hidden lynchpin on which the success of DG automation hinges.

So, it's best to think of data integration as a key infrastructure layer that enables many different forms of DG, whether directly or indirectly. Note that—as vendor tools evolve—infrastructure designed for data integration can also support data quality, metadata management, and other data management techniques that are core to DG. With that in mind, data integration infrastructure is a hidden lynchpin on which the success of DG automation hinges. In Figure 15, data integration is the layer through which all enterprise data—whether governed or not—flows. According to an interviewee: "Most of the data we govern closely is integrated into an enterprise data repository, then distributed to other systems. All that data and all of those transformations and verifications run through Informatica PowerCenter, which is critical to our data governance program."

Representative data integration tools come from Business Objects, Informatica, and SAS. These same vendors produce data integration suites that include tools for multiple forms of data integration (i.e., ETL, EII, replication), plus tools for practices related to data integration, like metadata management, master data management, data quality, and data profiling/monitoring. Over time, expect each of these suites to coalesce into unified platforms that seamlessly support all data integration and quality functions from a single console. Also, expect these platforms to fold in data governance functions, as DG becomes more closely associated with these tool types.[9]

Databases

Many of the users interviewed for this report have taken a physical approach to data governance, in that they relocate data that they want to govern into some kind of central database. In a lot of ways, this is a data warehousing mentality applied to data governance. The upside is that it's easier to assure the security and privacy of data—as well as perform data quality and master data management operations—when all governed data is in one place (or fewer places). The downside is that you have to integrate and synchronize large volumes of complex data.

Some users consolidate governed data into an enterprise data repository (EDR).

The resulting central database varies a lot. Sometimes it's a recognizable enterprise data warehouse (EDW), but more often it's a unique operational data store (ODS). Many of the users interviewed used the term *enterprise data repository* (EDR), which is a collection of databases, including data warehouses, marts, ODSs, and semantic repositories. In these cases, an EDR contains a mix of detailed source data, lightly summarized data, multidimensional cubes, metadata, master data, and even content (in the sense of natural language text).

An EDR may also subsume the metadata repositories mentioned earlier and similar databases, like a data glossary (38% in Figure 14). Almost all EDRs and other databases specifically for DG are homegrown designs built atop a database management system (23%).

From an architectural viewpoint, an EDR or other DG database can be a hub similar to the repository in the center of Figure 15. Or, the EDR could include a central hub, plus the databases and repositories of various integration applications, as seen on the right side of Figure 15.

Dedicated DG applications

By "dedicated," we mean that DG is the primary purpose of the application, as opposed to the secondary body of DG functions found in some tools. This kind of dedicated DG application is rare, at the moment, but TDWI anticipates that more will appear in 2008 and 2009 in response to the considerable demand among users.

While conducting interviews for this report, TDWI Research encountered a few SAP users with dedicated DG applications. This includes SAP for Governance, Risk, and Compliance (GRC), which is a framework that includes applications for access control, process control, global trade

[9] For a complete survey of data integration vendors and tools, see the TDWI Technology Market Report *Data Integration Tools* (Q4 2007), available to TDWI Members at www.tdwi.org/research.

services, risk management, and so on. Other interviewees were using dedicated DG applications and related services from BackOffice Associates, Inc., and IDS Sheer.

There are also numerous applications that have a substantial body of governance functions, despite an overall focus elsewhere. Examples include SAS Financial Management Solution (heavy on SOX and other compliance guidelines), SAS Drug Development Solution (guarantees compliance with HIPAA), and SAS Model Manager (addresses regulations around risk model management and validation). Then, there's the new SAP Information Lifecycle Management (ILM) solution, which reminds us that data needs governance as it ages and moves from active service to archive and all the steps in between.

Collaborative documents for DG

DG processes generate a lot of documents that need managing.

Like any other program, data governance generates a lot of paperwork through which people collaborate. This includes policy documents, forms for change requests and other procedures, board descriptions, documents chronicling board decisions, project timelines, meeting minutes, and a plague of other documents. DG board members and others can hurl these at each other via e-mail. But it's best to store them in a central place where everyone can access the latest version as they need it. The mechanism for sharing Microsoft Office files (which is what most of these are) can be as simple as a folder on a shared network drive or as complex as a full-blown content management system. Somewhere in the middle is the DG repository, whether it's a semantic data repository or an EDR. Most repositories are capable of storing Office files, and it makes sense to keep these documents in the same software system that automates the DG process.

Professional services for DG

DG consulting is hot right now because DG is new and unknown.

Tool types aside, consulting services are also an important component of DG solutions. Data governance is a hot topic for consultants today, because user organizations need a kick start to help them understand what DG is, what they would do with it, and how to get started. For example, the data management practice at Boston-based Collaborative Consulting has developed expertise and templates for DG engagements. And software vendors with DG-related tools offer relevant professional services, like the Data Governance Workshop from Trillium Software and the Data Governance Boot Camp from DataFlux.

USER STORY

Customers get the data they want, when they want it, from a DG-controlled, Internet-accessible database.

"Due to competitive pressure, Blue Cross and Blue Shield of Kansas City (BCBSKC) needed to leverage data as a strategic asset and share better information with its customers, including insured members, employer groups, and healthcare providers," said Darren Taylor, the vice president of information access at BCBSKC. "To achieve this goal, we adopted a strategy with three prongs—foundation, accountability, and empowerment." Building a solid *foundation* meant integrating data from dozens of sources into an enterprise data warehouse that provided a consistent view of subject area data, regardless of the source. *Accountability* for the data integration effort was created by forming a dedicated center of excellence focused on data management and business intelligence solutions. Defining and managing data access and data usage became simpler once data was centralized and a specific division was named accountable. Technical matters like data quality and consistent master data definitions also became easier to address as a result. The solution now *empowers* BCBSKC by providing business systems and processes with required data; by providing a dependable basis for decision making; and by providing constituents with their own Internet portals to access the consistent and secure data they want, when they want it.

Recommendations

Staff DG with a cross-functional team and a powerful executive sponsor.

- **Always remember that DG success is mostly about the four Ps.** People collaborate to create policies and procedures that constitute the DG process. In other words, most DG tasks are interpersonal and organizational, with certain data management technologies playing a supporting role.
- **Recognize that DG is inherently cross-functional and collaborative.** Accordingly, staff the DG board with diverse people from diverse departments, so all their needs are all represented. Staffing should also mix executives and workers across both business and IT.
- **Elect a board chairperson who will make DG a priority.** The chair should be a high-placed executive who both wields broad influence and evangelizes DG as an imperative. His/her executive mandate is required to overcome territorial issues and general organizational intransigence. As an alternative, consider co-chairs who together represent business and IT.
- **Don't practice data governance in a vacuum.** Coordinate it with related data management practices, data-intense business initiatives, and other forms of governance—especially IT governance and corporate governance.

Define appropriate forms and levels of control for DG.

- **Note that DG usually boils down to some form of control.** The catch is to know the different meanings of control (limiting data access, expanding data integration, controlling data quality) and to define the appropriate amount of control for each (strict governance versus loose guidance).
- **Balance the opposing goals of DG.** The most obvious conflict is between compliance goals that limit data access and business integration goals that expand data access. Other opposing goals to balance include business versus technology, data content versus data usage, and departmental versus enterprise data ownership.

Start with one of DG's three pillars.

- **Start with compliance, business transformation, or data integration.** These are the three pillars of DG, because they're the most common starting points, and therefore represent the most pressing goals of organizations. Note that some organizations start with technical issues like data quality or master data management; impact is greatest when these support higher business priorities, as described by the three pillars.
- **Think of DG as a transformational process.** In other words, change is a constant with DG, and change needs management. Don't get so deep into requirements for compliance, quality, and integration that you forget to enable all these with change management via a DG board.
- **Be aware of common intersections.** DG regularly coordinates with business initiatives and technical implementations. Knowledge of these can guide your plan for which to address and in which order. And the collection of initiatives and implementations you choose to address will determine the scope of your DG program and how it is staffed.

Tool support for DG has its place, and its place will grow.

- **Automate DG tasks with software when possible.** Although DG is largely about the four Ps, software automation can give it speed, accuracy, auditability, and scalability. Today, this is achieved mostly via selected functions in data management tools for data quality, integration, and metadata management. A few dedicated DG applications are available from vendors. Expect vendors to produce more tools and better automation for DG as users better define their tool requirements and DG becomes a priority in more organizations.

Business Objects, an SAP company

As an independent business unit within SAP, Business Objects transforms the way the world works by connecting people, information and businesses. Together with one of the industry's strongest and most diverse partner networks, the company delivers business performance optimization to customers worldwide across all major industries, including financial services, retail, consumer-packaged goods, healthcare and public sector. With open, heterogeneous applications in the areas of governance, risk and compliance; enterprise performance management; and business intelligence; and through global consulting and education services, Business Objects enables organizations of all sizes around the globe to close the loop between business strategy and execution.

www.businessobjects.com

Collaborative Consulting

Collaborative Consulting optimizes clients' business and technology capabilities. Our approach facilitates collaboration between business and technology groups to ensure the technology solution is designed to deliver business value. Collaborative uses exceptional innovation and creativity to help companies rein in their voluminous data and use it as a strategic asset. We offer end-to-end data management, from strategy through implementation. Additionally, we make sure that all data initiatives are aligned closely with clients' business objectives.

www.collaborativeconsulting.com

DataFlux

DataFlux enables organizations to analyze, improve and control their data through an integrated technology platform. With DataFlux enterprise data quality and data integration products, organizations can more effectively and efficiently build a unified view of customers, products, suppliers or any other corporate data asset. A wholly owned subsidiary of SAS (www.sas.com), DataFlux helps customers rapidly assess and improve problematic data, and build the foundation for enterprise data governance. Effective data governance delivers high-quality information that can fuel successful enterprise efforts such as risk management, operational efficiency and master data management (MDM).

www.dataflux.com

Exeros

Exeros, the leading data relationship discovery company, accelerates time to market for master data management, data governance and data integration initiatives. Exeros' customers include some of the world's largest financial and manufacturing organizations. The company's cross-system data analysis and validation software discovers and validates business rules, data lineage and unknown data inconsistencies within and between corporate data sources.

www.exeros.com

Informatica Corporation

Informatica Corporation delivers data integration and data quality software and services that help organizations gain optimum value from their information assets. Informatica's open, platform-neutral software accesses data of virtually all types and makes it accessible, meaningful, and usable to the people and processes that need it. With products that encourage collaboration across the enterprise, Informatica reduces costs, speeds time to results, and scales to handle data integration projects of any size or complexity. That is why Informatica is known as the data integration company.

www.informatica.com

SAP

SAP is the world's leading provider of business software. More than 41,200 customers in more than 120 countries run SAP® applications—from distinct solutions addressing the needs of small and midsize enterprises to suite offerings for global organizations. Powered by the SAP NetWeaver® platform to drive innovation and enable business change, SAP software helps enterprises of all sizes around the world improve customer relationships, enhance partner collaboration and create efficiencies across their supply chains and business operations. SAP solution portfolios support the unique business processes of more than 25 industries, including high tech, retail, financial services, healthcare and the public sector.

www.sap.com

SAS

SAS is the leader in business intelligence and analytical software and services. Customers at 44,000 sites use SAS software to improve performance through insight from data, resulting in faster, more accurate business decisions; more profitable relationships with customers and suppliers; compliance with governmental regulations; research breakthroughs; and better products and processes. Only SAS offers leading data integration, storage, analytics and business intelligence applications within a comprehensive enterprise intelligence platform. Since 1976, SAS has been giving customers around the world THE POWER TO KNOW®.

www.sas.com

Trillium Software

Harte-Hanks Trillium Software® has been selected by companies worldwide, both large and small, to improve their operational and analytic business decisions through accurate and timely information. Trillium Software offers an integrated suite of Total Data Quality software and services architected to discover and correct today's data quality problems and establish a platform prepared for tomorrow's yet unknown data challenges. The Trillium Software System® is recognized as critical to the success of customer relationship management, master data management, customer data integration, data warehouse, business intelligence, enterprise resource planning, supply chain management, e-business, and other enterprise applications, and data integration, data migration, data stewardship, and data governance initiatives.

www.trilliumsoftware.com

TDWI RESEARCH

TDWI Research provides research and advice for BI professionals worldwide. TDWI Research focuses exclusively on BI/DW issues and teams up with industry practitioners to deliver both broad and deep understanding of the business and technical issues surrounding the deployment of business intelligence and data warehousing solutions. TDWI Research offers reports, commentary, and inquiry services via a worldwide Membership program and provides custom research, benchmarking, and strategic planning services to user and vendor organizations.

1201 Monster Road SW
Suite 250
Renton, WA 98057

T 425.277.9126
F 425.687.2842
E info@tdwi.org

www.tdwi.org