TDWI

MONOGRAPH SERIES

MARCH 2009

Enterprise Information Management: In Support of Operational, Analytic, and Governance Initiatives

By Philip Russom

Senior Manager, TDWI Research

The Data Warehousing Institute

SPONSORED BY

tdwi

THE DATA WAREHOUSING INSTITUTE

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, awards, and events. Before joining TDWI in 2005, Russom was an industry analyst covering BI at Forrester Research, Giga Information Group, and Hurwitz Group. He's also run 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 Our Sponsor

SAP is the world's leading provider of business software, offering applications and services that enable companies of all sizes and in more than 25 industries to become best-run businesses. SAP has more than 82,000 customers in over 120 countries. The SAP® BusinessObjects™ portfolio transforms the way the world works by connecting people, information and businesses. With open, heterogeneous solutions in the areas of business intelligence; information management; governance, risk and compliance; and enterprise performance management, the SAP BusinessObjects portfolio enables organizations to close the gap between business strategy and execution. For more information, visit www.sap.com.

Defining Enterprise Information Management

In most organizations today, data and other information is managed in isolated silos by independent teams using diverse information management tools for data integration, quality, profiling, federation, meta- and master data management, and so on. However, there's a trend toward enterprise information management (EIM), a practice that holistically coordinates teams and integrates tools. Through team collaboration and tool interoperability, EIM seeks to improve the "four Cs," namely the completeness, cleanliness, consistency, and currency of structured and unstructured data.

The four Cs are worthy goals from a technology viewpoint, and they certainly prepare data for the next step, which is to share and leverage information across multiple business units of an organization and with business or trading partners. Once data is ready to be shared as an organizational asset this way, however, the ultimate goal of EIM becomes to achieve strategic, data-driven business objectives, like fully informed operational excellence and business intelligence, plus related goals in governance, risk, and compliance. All these steps combine to form an iterative process for EIM. (See Figure 1.)

Figure 1. An iterative process for achieving enterprise information management.

With all that in mind, here is TDWI's nutshell definition of enterprise information management:

> EIM is a best practice for creating, managing, sharing, and leveraging information in an enterprise, holistic manner that's aligned with strategic, data-driven business objectives.

Given this basic definition, certain corollaries follow:

- **EIM depends on data management tools and techniques.** EIM is not an independent tool type. Instead, EIM unites related information management tools and their best practices into a unified practice and infrastructure. Organizations use information management tools and techniques in implementing their EIM strategies, and organizations need these to achieve their EIM goals.
- **EIM is a matter of degree.** Most organizations already practice unconnected pieces of EIM at a minimal level, although they may call it something else. True EIM is a well-thought-out strategy of orchestrated phases (as seen in Figure 1). Iteratively working through the EIM process increases the degree to which it is practiced.

EIM technology practices must be holistic and aligned with business strategy.

Furthermore, enterprise information management has two important goals:

- **EIM must unite diverse information management practices.** This is mostly about coordinating the development efforts of data management teams and enabling greater interoperability among their servers. There are different ways to describe the resulting practice, and users who've achieved EIM call it a holistic, coordinated, integrated, or unified practice. Regardless of the adjective you prefer, the point is that EIM practices must be

inherently holistic, if you're to improve and leverage data on a broad enterprise scale. Likewise, you should assume that true EIM is always holistic by nature.

- **EIM must support strategic business objectives.** In particular, EIM must support data-driven initiatives like operations, analytics, and governance. Early in the EIM process, this is mostly about defining business objectives and related technology requirements for unifying EIM. Later in the process, this is about executing a business plan that leverages EIM to achieve the objectives. When EIM supports strategic goals, EIM itself becomes strategic.

SURVEY SAYS: 45% of organizations surveyed already coordinate diverse information management practices.

TDWI Research recently asked TDWI conference attendees a few questions about EIM, and found that diverse information management practices are moderately coordinated today. Cross-team coordination and collaboration are base requirements for EIM, and 45% of survey respondents report already doing this on a moderate level. (See Figure 2.) At the other end of the spectrum, only 9% of respondents work in organizations with very low coordination.

With EIM in mind, rate the level of formal coordination of the information management practices in your organization.

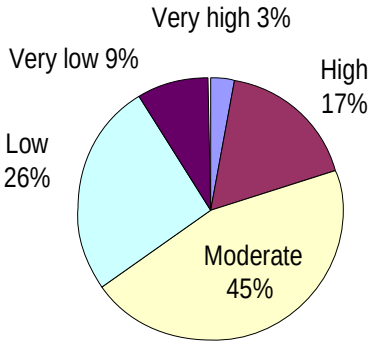

Figure 2. Based on 112 responses, Feb. 2009.

What it means for EIM to be holistic and strategic

Making EIM holistic and strategic are important goals, as described above. But getting there demands a number of adjustments to current technology solutions and business processes. For example, truly holistic enterprise information management has many organizational and technical requirements.

EIM is inherently collaborative and cross-functional.

Team collaboration. Let's recall that many organizations manage data in isolated silos strewn across the enterprise, using a variety of tools and teams. A certain amount of coordination among these teams is inevitable, because their solutions interact. For example, data integration tools regularly call data quality tools, data federation tools connect to operational databases, and data quality and master data management solutions often require changes to other, related data management solutions. Although this is a good start, EIM demands much more. Ideally, the combination of all data management solutions involved should form a recognizable architecture—even if it's loosely federated—that's held together by agreed-upon methods for the deep tool and platform interoperability that EIM assumes. Getting to this point usually demands the creation of a new, broader team structure, like a central competency center, a data stewardship program, or a data governance board.

Cross-functional. This has many meanings. An early implication of EIM is the coordination of multiple data management teams, each with a different discipline. On the business side, EIM demands that multiple business units coordinate data-driven, cross-unit business processes. Though EIM looks like a technology practice, its purpose is to enable business practices, thus providing an opportunity for business-to-IT alignment, which is also cross functional.

Interoperability. Acquiring a long list of tool types for an organization's EIM portfolio is important, perhaps even a base requirement. Equally important is that each piece of the portfolio

integrates and interoperates with others appropriately. To set the proper expectation, don't assume 100% interoperability; tool integration should be selective, based on prominent business and technology needs.

Data sharing. This is one of EIM's loftiest goals, and it's a prerequisite for data-driven business objectives. In this regard, EIM must do two things. First, data should be managed in a way that prepares it for sharing. At the base level, this entails improvements to data's quality and semantics, plus the adoption or development of data standards, connectivity methods, and usage policies. Second, EIM should provide an infrastructure through which data can be shared broadly.

Enterprise scope. EIM is inherently broad, involving some definition of "enterprise scope." But don't start by addressing the entire enterprise. That would be a risky "big bang" project. Some organizations begin by deepening the integration and interoperability of the information management tools in their existing BI and data warehousing infrastructure. Others do the same with platforms for ERP or CRM. Still others begin with EIM for financials or the supply chain. As diverse pockets of EIM develop from these starting points, the next step is to connect the dots and unify them at a depth that is appropriate to enterprise goals.

For EIM to be strategic, it must align with the business by explicitly supporting corporate objectives of various types:

EIM supports business objectives, cross-functional consistency, and the sharing of enterprise data assets.

Enterprise information management supports corporate objectives. Sure, there are technology goals to support, but the base requirement for making EIM strategic is to support business goals. EIM provides an information infrastructure for data-driven business initiatives like global spend analysis, 360-degree customer relationship management, data as an enterprise asset, and operational business intelligence.

Enterprise information management yields cross-functional consistency. Again, the technology provides a certain amount of consistency due to common definitions of business entities expressed through shared metadata and master data. Strategic business goals that reach across business units and other functions then become attainable—goals like accurately identifying customers, uninterrupted business processes that span multiple applications and departments, faster financial closings, and business intelligence based on enterprise visibility.

Enterprise information management enables the enterprise to pool and share resources. This is key to the success of many business initiatives. For example, EIM gives business intelligence an enterprise breadth and depth of visibility that's otherwise difficult to attain, thus enabling views into corporate performance appropriate to strategic, tactical, and operational BI. Similarly, EIM provides data synchronization and consistent semantics for heavily distributed operational applications (typically for ERP, CRM, and financials), thus helping unite otherwise disparate business units and cross-unit processes.

Once these components of EIM are in place, other strategic goals fall into line:

Enterprise information management helps IT align with business. And some organizations need all the help they can get with this strategic goal.

Enterprise information management enhances auditability, which makes the enterprise more compliant. Many corporations are struggling with these strategic goals. A good EIM program will improve the quality of master, reference, and meta data, plus tighten policies for data use, traceability, security, and privacy. Combined, these improvements make data's lineage, impact, and usage easy to corroborate in audit and compliance situations.

Enterprise information management's support for business goals makes its ROI apparent. When EIM reaches beyond technology to support business objectives that have tangible returns—for revenue, profit, cost reduction, and share value—a portion of those returns are easily associated with EIM.

SURVEY SAYS:
46% of organizations surveyed said EIM could be highly strategic.

TDWI Research recently asked TDWI conference attendees how strategic they think a holistic approach to EIM could be in their organizations. Forty-six percent of survey respondents reported that it could be highly strategic (see Figure 3), whereas an additional 20% felt it could be *very* highly strategic. Few survey respondents said that EIM is not very strategic (8%) or not strategic at all (1%). This indicates that EIM has a strong potential for strategic impact.

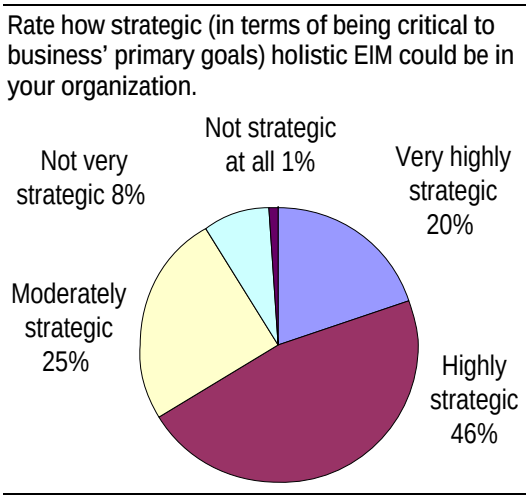

Figure 3. Based on 110 responses, Feb. 2009.

Examples of information management tools and practices being integrated in EIM

EIM focuses on data integration and related practices.

Just about any information management tool or practice can be involved with EIM, as can various applications, databases, file systems, and so on. However, EIM today centers around data integration and closely related tools and practices for data quality, data semantics, MDM, and content management.

Data integration. EIM depends heavily on data integration for sharing data across multiple business units. This is true whether the form of data integration is extract, transform, and load (ETL), data federation, replication, or hand-coding. Furthermore, data integration platforms also provide other EIM necessities, like support for a wide array of APIs, gateways, data standards, and other interfaces, including Web services and SOA. Many users prefer the master data and metadata management capabilities of their data integration platforms, and some of these platforms include functions for data quality, profiling, and monitoring. Put all this together, and you see that a data integration platform is essential infrastructure for EIM and, therefore, a starting point for many users. Even so, the data integration platform should also integrate tightly with application integration platforms to give EIM the broad reach it needs.

Data quality. Data integration and data quality techniques have become closely related in recent years, to the point that the two are regularly folded together into a seamless solution. The most important task in preparing data for EIM's broad sharing is to improve the quality of data's content and structure. Data quality work includes data profiling (to study data in the early phases of solution development) and data monitoring (which re-profiles data as the deployed data quality solution operates on it daily, to assure that quality metrics are being met).

Metadata management. Another important task in preparing enterprise data for EIM's sharing is to improve various types of semantic data. This is a holistic exercise, since the goal is to ensure that descriptions of data are complete and meaningful, as well as consistent when multiple applications and databases handle common data. Most organizations begin addressing these goals by standardizing metadata and similar "data about data" found in the metadata repositories, data

dictionaries, glossaries, and semantic layers of various data management tools, operational systems, and BI tools for reporting and analysis.

Master data management (MDM). Applications of varying types are more interdependent today than ever before. Through data integration and application integration technologies, applications share growing volumes of information to achieve the business objectives associated with EIM. Due to the complexity of the situation, it can be difficult to find the correct or best sources of data and map them accurately between applications. For this reason, many organizations have recently embraced new approaches to MDM. The modern practice of MDM enables users to consistently define data and how it should be used across various types of applications and their business units, whether in the context of operational processes or analytic decision making. And MDM can make data more governable and auditable, as well. In other words, MDM enables many EIM goals.

Data warehousing. One of the direct beneficiaries of EIM is the enterprise data warehouse (EDW). After all, analyses, reports, and decisions based on the EDW are only as good as the data managed within it. And EIM greatly improves the quality, semantics, and currency of the EDW's data. The benefits of EIM likewise extend to databases similar to the EDW, like data marts, cubes, and operational data stores, as well as many operational databases.

Content management. A lot of information useful for EIM practices originates in documents of mostly text, like word processing files, Web pages, content management systems, and so on. Text analytics tools and other technologies that employ natural language processing (NLP) can convert this information into structured data that EIM and similar practices can use.

Now that we've defined EIM and shown its holistic and strategic possibilities, let's look at the ramifications for three areas of special significance, namely operational systems and business processes, business intelligence and data warehousing (BI/DW), and governance and similar organizational structures.

EIM for Operational Systems and Business Processes

Operational excellence requires excellent information.

Operational data evolves in the context of a business process, but EIM still applies.

In operational and transactional applications, data usually exists and evolves in the context of a business process. As the process progresses from one state to another, information about the business entities involved in the process changes. For example, as a claim moves through an insurance company's claim process, the computerized claim record grows and evolves as the state of the claim changes. Operational data evolves with loans in financial services, patients in healthcare, bills of material in manufacturing, or procurement orders in any industry. Some complex processes—like fulfillment or supply chain—unfold across multiple departments and include diverse steps and branching for routing, review, and approval. This makes the continuous updating of operational data even more complicated.

"Data in motion"—as seen in these examples of process-driven operational applications—is quite different from the "data at rest" that's typical of BI/DW. Yet, EIM has benefits for data in motion, just as it does for data at rest. In particular, some of the technologies associated with EIM can make operational data more useful at each stage of the process by increasing consistency across states and by improving accessibility for diverse departments. All these enhance the business process (not just data and applications) such that EIM contributes to operational excellence.

EIM satisfies the four Cs required for operational excellence.

EIM helps operational data become complete, clean, consistent, and current.

Operational applications and business processes are complex systems that have many data requirements. But EIM helps satisfy some of these, as seen in the four Cs:

- **Complete information.** As pointed out earlier, business people lower in the org chart make a lot of operational decisions every day, and these contribute to operational excellence when they're enlightened with complete, clean, consistent, and current information. For example, for the sake of good customer service, employees and applications at every customer touch point need complete and current customer information. Likewise, accurate fulfillment, shipping, billing, and manufacturing floor management processes need complete information about products and supplies. Since these departments may have their own applications, your EIM strategy should seek to disseminate relevant information for cross-department processes, whether multiple types of applications or multiple instances of the same one are involved. As you can see, EIM helps with the classic problem of siloed applications.

  Again, integration infrastructure is a key enabler for EIM, whether it focuses on operational data integration, enterprise application integration, or a mix of both. Web services and SOA are popular with operational applications (more so than with BI/DW), so they are likely enablers for EIM in support of operational excellence.

- **Clean information.** Operational applications are notorious as the origin of most data quality problems. In the spirit of EIM, these applications need data quality functionality to cleanse, standardize, and enhance data. These functions are particularly effective when applied "upstream" just before new or changed data is committed to an application's database. Note that going upstream usually requires data quality tools that can operate in real time. Downstream data quality functions applied to application databases—like match-and-merge and deduplication—are useful for reducing redundant records, which are inevitable in most operational applications.

- **Consistent information.** Given that many people and departments will work on a business process, it's sometimes difficult for business users and applications to identify and match customers, products, and other business entities with any speed or accuracy. For example, the modern enterprise wants to recognize a customer when he/she returns or touches a different business within the enterprise. Likewise, every link in an efficient supply chain will identify parts and supplies accurately, perhaps even match those that are equivalent commodities. In the EIM paradigm, these goals are achieved through consistent information definitions expressed via metadata, master data, and the semantic layers of tools, applications, and platforms.

- **Current information.** Some business processes and practices need to operate at or near real time, as with just-in-time inventory, build-to-order manufacturing, utility grid monitoring, and customer up-selling and cross-selling. In a similar vein, operational BI depends on application data that's integrated quickly or frequently. These practices handle time-sensitive or highly changeable information, which goes stale or loses importance quickly after it is integrated or synchronized. So, for information of extreme currency, it makes sense to integrate operational data on the fly through near-real-time technologies like data federation, EAI, and services. In the EIM paradigm, real-time technologies interoperate with many other technologies—including applications, databases, and a variety of integration tools—and that demands coordination in development and interoperability in deployment.

Technical attributes of EIM for operational excellence

EIM demands adjustments in operational applications and related infrastructure.

Don't be surprised if the EIM practices discussed here for operational applications and business processes sound like something you're already doing. EIM for operational excellence is, in many ways, a matter of degree. In other words, EIM involves doing more of what you're already doing in technical areas (like integration and interoperability) and business areas (like IT-to-business alignment and supporting strategic business goals). Here are some adjustments to current practice that you should expect if you plan to reach the next level of data management via EIM.

Extend enterprise application integration (EAI) infrastructure. Most EAI deployments transport small messages among a few applications. EIM practices, being more data-oriented, force EAI to carry larger messages and to transform data in more complex ways. EIM may also lead to using EAI infrastructure more like an enterprise bus, replete with data-oriented services. For instance, EIM practices lead to more real-time services for data quality and operational BI. EIM also leads to more interoperability among integration servers, so expect EAI to interface more with ETL and other data integration tools.

Apply data quality (DQ) functions to operational applications upstream. Although any good application will have data validation and standardization routines that parse data typed into its user interface, upstream DQ functions take this to a new level that will fix many data quality problems and deficiencies. This requires DQ tools that can respond in real time to any type of data, and you'll probably embed the same DQ function in multiple applications. So, consider retrofitting these functions to preexisting applications through a single set of DQ services.

Beef up master data management (MDM). For cross-system consistency of data definitions and data usage—a primary goal of EIM—MDM will give you even greater results than similar efforts in metadata. This is especially true of extreme cases, like business processes that flow across multiple business units and their applications.

Enlighten EIM with information lifecycle management (ILM). TDWI Research defines ILM as a best practice that applies technology policies and business rules to managing information as its maturity and usefulness evolve over time. The assumption is that data's business value and technology requirements change as it matures and ages. So information should be managed by IT and leveraged by the business in different ways as the information wends its way through various stages of its lifecycle. In general, ILM focuses on policies for late-lifecycle stages, like data retention, archiving, and final disposition. Hence, ILM can give EIM a method for handling mature data in alignment with its unique technology requirements and business value.

Share and leverage information through operational data integration. TDWI Research defines operational data integration (OpDI) as "the exchange of data among operational applications, whether in one enterprise or across multiple ones."[1] Like EIM, OpDI addresses real-world problems and opportunities with operational data. And OpDI projects involve multiple data management practices, plus coordination with business managers. Furthermore, OpDI relates to ILM in that OpDI techniques are often applied to mature data in its late lifecycle stages. Here are examples of common OpDI projects:

- **Data migration.** Data migrations typically abandon an old platform in favor of a new one, as when migrating data from a legacy hierarchical database platform to a modern relational one.

[1] For more information, see the TDWI Best Practices Report *Operational Data Integration: A New Frontier for Data Management*, available for free download at www.tdwi.org/research/reportseries.

Sometimes the abandoned database platform isn't really a "legacy"; it simply isn't the corporate standard.

- **Data consolidation.** For example, many organizations have multiple customer databases, which need consolidation to gain a single view of customers. Data mart consolidation is a common example in the BI world. And consolidating multiple instances of a packaged application into one involves consolidating the databases of the instances.
- **Database upgrade.** Upgrading a packaged application for ERP or CRM can be complex when users have customized the application and its database. Likewise, upgrading a database management system is complicated when users are two or more versions behind.
- **Database collocation.** This is often a first step preceding other data migration or consolidation project types. For example, you might collocate several data marts in the enterprise data warehouse before consolidating them into the warehouse data model. In a merger-and-acquisition, data from the acquired company may be collocated with that of the acquiring company before consolidation.
- **Data synchronization.** When IT systems share data in common—typically about business entities like customers, products, or financials—it may be necessary to synchronize data across the redundant systems so the view of these business entities is the same from each application and its database. For example, data synchronization regularly syncs customer data across multiple CRM and CDI solutions, and it syncs a wide range of operational data across ERP applications and instances.
- **Business-to-business (B2B) data exchange.** For decades now, partnering businesses have exchanged data with each other, whether the partners are independent firms or business units of the same enterprise. B2B data exchange is a mission-critical application in industries that depend on an active supply chain that shares a lot of product information, such as manufacturing and retail. It's also critical to industries that share information about people and money, like financials and insurance.

USER STORY EIM concepts can be applied on a global scale.

"We have one global instance of SAP, and you wouldn't believe the volume and complexity of data flowing in and out of it," said a data architect at a multi-national consumer packaged goods firm. "For one thing, we sync large volumes of data across geographic regions. For another, we have to convert a lot of this data, because most geographies have region-specific accounting standards and currencies. It helps that our chart of accounts, products, and other important business entities are defined as master data in SAP, and the master data is published outbound from corporate headquarters. We likewise enforce master data and metadata standards with non-SAP systems, especially tools for data integration, application integration, and business intelligence. This helps us keep information consistent, as we exchange data through two large infrastructures: one for enterprise application integration, the other for data integration, mostly ETL. Today, this is 'spaghetti architecture,' but we're currently re-architecting it into a series of regional hubs, where most communication will be through services. We founded an integration competency center to coordinate the work. Our eventual goal is to have one enterprise-scale information management infrastructure that's built from various integration technologies."

EIM for Analytics, Business Intelligence, and Data Warehousing

Great decisions require great information.

BI and DW are fact-based methods, and EIM improves the facts.

EIM yields accurate facts, which leads to accurate decisions.

Business intelligence (BI) is a fact-based methodology that assists with decision making on multiple organizational levels. Relevant to EIM, decisions are based on facts, and the facts are defined by data. The numeric content of the average BI report attests to this. Even when the numbers are expressed visually in charts, visualizations, dashboards, and scorecards, the numbers are still there defining facts. As another example, consider the fact table. It's a common data structure in data warehousing (DW), and it models an organization with data to reveal facts about organizational performance. Hence, business people consume the reports and data analyses of BI/DW so they can make fully informed decisions, based on data-defined facts about the business. EIM's role is to assure that data leads to accurate facts, which in turn lead to accurate decisions.

EIM contributes to the three standard practices of decision making.

BI's dependence on data-defined facts is fundamental, and the dependence is apparent with all types of decision making. For example, think of the average organizational chart: it has executives and line-of-business managers at the top, middle managers in the middle, and various knowledge workers and operational workers toward the bottom. Although the lines between these three user constituencies are fuzzy and movable, they fairly accurately align with the three standard practices for decision making, namely strategic, tactical, and operational BI. Of course, BI also reaches laterally across multiple departments.

Nowadays, everyone makes decisions, so everyone benefits from EIM.

Hence, a mature BI and DW implementation provides reports, analyses, and data delivery mechanisms that address multiple decision-making practices at multiple levels of an organization. Given EIM's role in improving facts, EIM contributes to the decision-making work of a wide variety of people and departments, because so many of them consume BI and DW information that's fed through EIM. As discussed earlier, when a technology has enterprise-scale influence—as EIM clearly does in the realm of BI/DW—it should be considered strategic infrastructure and fostered to ensure that it continues to contribute to strategic business goals.

EIM satisfies the four Cs required of BI/DW.

The four Cs are the primary goals of all forms of EIM.

BI and DW are complex practices that have many data requirements. But EIM helps satisfy the top requirements, as seen in the four Cs:

- **Complete information.** Most of the facts, business entities, and performance metrics represented in a data warehouse or BI report are based on data that came from multiple enterprise applications and third-party providers. The data consolidation, enrichment, and sharing practices and data integration functions of EIM ensure that decisions are based on a single, complete view of the facts.
- **Clean information.** One of EIM's most appealing value propositions is its ability to correct, standardize, enhance, and consolidate data coming from multiple sources. These and other data quality tasks should be integral components of EIM, else the quality of decisions suffers.
- **Consistent information.** With decision data coming from multiple sources, it can be hard to know where it came from (which is key to auditability) and how it represents a business entity (which is key to finding the best source for a specific report). These and other consistency

issues are resolved through the EIM disciplines of metadata, master data, and other semantic data.

- **Current information.** The three standard decision-making practices (and other BI practices, too) have varying requirements for information currency, defined as how recently data in a DW or BI report was refreshed. EIM handles decision data from many sources, but also processes it at right-time speeds (from monthly to milliseconds), as determined by the time-sensitivity of the data or a dependent decision-making process.

Technical attributes of EIM for BI/DW

The complex BI/DW technology stack is a model for EIM.

Integrating BI/DW tools is both a model and a starting point for EIM.

The technology stack for BI and DW is rather complex, including many tools, technologies, and platforms, plus a wide range of interfaces. This complexity is usually organized into five broad areas by related functionality.

- **Source systems.** Databases of operational and transactional enterprise applications, plus data from business partners and third-party providers.
- **Data integration infrastructure.** Data integration in the form of ETL, federation, replication, hand coding, and so on; plus closely related tools for data quality, data profiling, connectivity, and text analytics.
- **Data semantics**. Master data management (MDM) and metadata management; plus, similar "data about data" found in the metadata repositories, data dictionaries, glossaries, and semantic layers of various data management tools, databases, operational systems, and BI tools for reporting and analysis.
- **BI data stores.** Enterprise data warehouses (EDW), data marts, operational data stores (ODSs), data staging areas, cubes, spreadsheets, and other target databases; plus in-memory, appliance, accelerator, and flat-file versions of these.
- **Enterprise BI environment.** Tools for reports, analyses (whether based on OLAP or predictive analytics), dashboards, scorecards, charting, visualization, BI portals, and so on.

As you can see, the BI/DW technology stack involves a portfolio of tools and practices that we associate with EIM. In this way, it provides a precedent for applying many data management practices and related practices to the service of a common goal. In fact, many organizations begin their EIM efforts in the domain of BI, and then apply what they learned elsewhere. Hence, BI is a great place to start, because BI serves as a model for EIM and because BI needs it to become holistic and strategic.

BI practices need a few adjustments to achieve EIM.

BI/DW needs EIM to become holistic and strategic.

Although the average BI/DW technology stack involves a tool portfolio similar to that of EIM's, BI is not automatically practiced in a holistic or strategic manner. The bad news is that the four areas of the BI/DW technology stack are specialties that need coordination. The good news is that—if we ignore source systems—the other three areas of the stack are often implemented and maintained by a single BI team under a single BI director. So BI practices just need a few adjustments to attain the coordination of EIM. Most of these adjustments are a matter of extending or deepening practices that the BI team is already doing:

Extend BI's data integration (DI) infrastructure. BI and DW have a long tradition of ETL excellence. Less mature organizations need to extend DI beyond ETL to encompass other DI approaches—especially federation, replication, and services—to satisfy advanced EIM requirements for broad connectivity, operational BI, and real time.

Add more data quality (DQ) functions. Although basic DQ functions like name-and-address cleansing and standardization are common, most BI and DW implementations lack advanced DQ functions for match-and-merge, real-time cleansing, and data enhancement (e.g., appending third-party consumer data to customer records in the data warehouse).

Deepen tool interoperability. Despite the diversity of tools used by a BI team, integration and interoperability among tools is still minimal. Today, servers for DI and DQ call each other's APIs on occasion, whereas all servers involved with EIM should interoperate more deeply, preferably through Web services and SOA. In a similar issue, developers of diverse tools need visibility (and possibly the chance to share or borrow development objects) for the sake of developer collaboration, productivity, and object reuse. Achieving deeper interoperability, integration, and collaboration is easiest with a tool set that manages data integration, data quality, and related processes in a seamless development and deployment environment.

Broaden semantic data management. Metadata is the golden thread that stitches the complex BI/DW technology stack together, so sophisticated metadata management is fairly common in BI, especially if it provides full data lineage to the source and impact analysis to any report. In the spirit of EIM, however, BI professionals need to beef up master data management (MDM), plus coordinate metadata and master data with the semantic layers of reporting and analysis tools.

USER STORY Full-blown EIM crosses the line between operational and analytic applications.

"We integrate data from multiple internal applications, business partners, and third-party consumer data providers," said a business analyst in the financial group of a computer manufacturer. "Some of the data goes into operational applications for order entry and billing, plus a real-time decision engine that controls whether a customer gets financing. Other data goes into operational reporting and performance dashboards for executives; all these are refreshed daily or on demand. And then, most of the data is repurposed for analytic applications for risk management, debt delinquency rates, and monitoring our reserve balance, plus ad hoc analyses by business analysts. This whole mix of operational and analytic applications is held together by an enterprise information management infrastructure that's based on BI and data warehousing techniques and coordinated through a competency center."

The Role of Governance in Enterprise Information Management

No doubt you've noticed by now that enterprise information management (EIM) demands that technical and business people change how they manage and leverage data. There are numerous points of change, but at a higher level they all have a common goal of unification, whether that involves information systems interoperating more holistically or business units sharing more data strategically. The goal of unification is worth pursuing, but getting there involves a lot of change that must be managed. Furthermore, many of the changes demand collaboration and coordination on multiple levels, including multiple technical teams, multiple business units, and the alignment of both camps. At this level of complexity, the challenge reaches beyond technology and business—EIM is an organizational challenge.

EIM requires a lot of change, and data governance manages change very well.

DG gives EIM procedures for change management, among other things.

Many organizations are addressing the coordination challenges of EIM and similar practices by creating organizational structures for various kinds of governance. These may be relatively high-level, like corporate governance or IT governance. However, data governance (DG) is a specific form of governance that's most relevant to EIM. Here's a comprehensive definition of data governance from TDWI Research:

> 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.[2]

DG and EIM have common goals, but are complementary in approach.

Obviously, DG and EIM have common goals, namely to improve the four Cs of data. With both DG and EIM, the point of these improvements is to make the sharing of data easier, quicker, and broader. Furthermore, both DG and EIM should be aligned with business to support strategic business objectives.

DG gives EIM process; EIM gives DG vision.

Yet, DG and EIM also have significant differences. Data governance establishes and enforces policies and procedures, whereas EIM is a strategy for managing information technology so that it will comply with those policies and procedures. EIM, due to its holistic nature, aims for unified data management tools and practices, whereas DG's bureaucratic nature tends to deal with one data management practice, data source, or information issue at a time. EIM is a somewhat fuzzy concept that's more of a destination than a journey, whereas DG is all about defining the policies and procedures of the journey. DG's bureaucracy can slow it down, whereas EIM's entrepreneurial spirit is often expressed through small and quick opportunistic projects.

Despite their overlap and differences, DG and EIM complement and complete each other. DG gives EIM the organizational structure, data-usage policies, and procedures for change management that it lacks. EIM gives DG a guiding vision and excitement that it all too often lacks.

DG and EIM both need cross-functional and collaborative teams.

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

For DG to be effective, it needs an organizational structure that's cross-functional, collaborative, and data oriented. The most obvious one is the data governance board (or committee). But there are other well-established types of organizational structures that can provide a precedent and a model for DG—and EIM, too. These include BI and data warehousing teams, data quality and stewardship programs, enterprise data architecture groups, and competency centers (sometimes called centers of excellence). Anyone planning a data governance program or EIM solution should learn from these, mimic their structures, and possibly borrow their resources.

[2] For a detailed study of data governance, see the 2008 TDWI Best Practices Report *Data Governance Strategies*, available online at www.tdwi.org/research/reportseries.

Data governance can be a critical success factor for EIM.

Any organization contemplating an implementation of the EIM concept should consider giving it structure through data governance or a similar organizational structure. Many firms found a data governance program before attempting information management tools and practices that require cross-functional coordination and change, as is the case with technical implementations for data quality, metadata management, and master data management (MDM).[3] EIM has similar cross-functional requirements, and so a data governance program can be a critical success factor for holistic and strategic EIM.

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."

Recommendations

EIM's primary goals are to unify data management practices and support business objectives.

Unify information management tools and practices through enterprise information management (EIM). Unification concerns three large tasks, namely coordinating design and development data management solutions, increasing integration and interoperability of deployed servers, and improving data to make more easily shared. Achieving these technology goals makes EIM comprehensive and effective.

Support business objectives through EIM. Align IT activities to strategic business objectives via EIM. After all, that's the ultimate goal of EIM. Supporting business goals this way makes EIM strategic. And it makes EIM's ROI more apparent.

All EIM efforts must address the four Cs.

Strive to improve the four Cs. EIM coordinates multiple data management practices to improve data's completeness, cleanliness, consistency, and currency. This is true whether EIM is applied to attaining fully informed BI or operational excellence.

Plan EIM with long-term goals in mind. EIM is a matter of degree. Start with a vision. Begin implementation of the vision within a well-bounded data management practice (like BI/DW, ERP, or CRM) where EIM is already practiced to a degree. Increase the degree of developer coordination and technology interoperability. Expect multiple EIM pockets to coexist. Connect these later. Get an executive sponsor who understands the value and competitive advantage of information.

---

[3] 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*. Likewise, the relationship between MDM and data governance is explained in the TDWI Best Practices Report *Master Data Management: Consensus-Driven Data Definitions for Cross-Application Consistency*. Both reports are available online at www.tdwi.org/research/reportseries.

EIM success depends on integration infrastructure.

Rely primarily on data integration and closely related data management practices. EIM includes multiple forms of data integration, like ETL, federation, replication, and synchronization. And it involves practices that go hand-in-hand with data integration, like data quality, data profiling, metadata management, master data management, Web services, SOA, and so on.

Extend integration infrastructure. Embrace multiple forms of data integration, plus related practices (especially data quality). Expect data integration infrastructure to interoperate with application integration infrastructure. Extend integration through services, when possible.

Improve data, but also semantics about the data.

Compile information about information. Know the data you're sharing, its origins, definitions, transformations, and impact. Document metadata, master data, and other semantic data in an open repository, so it can be shared. Include data lineage information that will help you prepare for audits. Enforce semantic data as a standard for cross-system and cross-department consistency.

Make BI more holistic and strategic with help from EIM. BI implementations of any maturity already have tools for data integration and other disciplines associated with EIM. Although these tools operate on the same data sets, they are not otherwise coordinated holistically, which EIM can cure. And EIM takes the data sharing of BI to the next level, making BI even more strategic.

Pursue operational excellence by improving operational data via EIM. Cross-department operational processes need the data consistency that EIM can provide though improvements to data semantics. Many forms of data quality make a similar contribution.

Don't forget that information lifecycle management (ILM) is part of EIM. After all, information has special information management requirements as it matures and changes business value. Depend on operational data integration for tasks related to ILM, like migrating and consolidating data as its lifecycle stages demand.

EIM is ruthlessly cross-functional, so it needs data governance for success.

Recognize that EIM is inherently cross-functional and collaborative. Attaining EIM's unification of information management tools and practices involves a lot of change that must be managed. Most of the changes demand collaboration and coordination at various levels, including multiple technical teams, multiple business units, and the cross-functional alignment of both of these. To be effective in this complexity, EIM needs data governance.

Put a data governance (DG) committee in place to ensure the ongoing success of EIM. DG and EIM complement and complete each other. DG gives EIM the organizational structure, data-usage policies, and procedures for change management that it lacks. EIM gives DG a guiding vision and mission that it sometimes lacks.