# **Recent Advances in Data Management**

ZELJKO PANIAN The Graduate School for Economics and Business University of Zagreb J.F. Kennedy Sq. 6 CROATIA zpanian@efzg.hr http://www.efzg.hr/zpanian

*Abstract:* - To increase the value of data as a business asset, companies and government organizations need to establish standards, policies, and processes for the usage, development, and management of data, to create the right organizational structure, and to develop the supporting technology infrastructure. As a practice with roots in corporate and information technology (IT) governance, data governance emerged. Data governance can be defined as the processes, policies, standards, organization, and technologies required to manage and ensure the availability, accessibility, quality, consistency, auditability, and security of data in an organization.

*Key-Words:* - Information technology (IT), data management, data integration, corporate governance, IT governance, data governance, compliance

# 1 Introduction

Although its value is not represented on the balance sheet, data is one of the most important assets in an organization. Data represents an organization's customers, employees, and suppliers; its activities and transactions; and its outcomes and results. Managed correctly, data can become an organization's most valuable asset, helping it to remain competitive and agile, to proactively meet customer needs, and to keep costs in check.

Companies and government organizations of all sizes are striving to manage data as an enterprise asset, to be shared and reused across multiple software applications and systems, business processes, and users throughout the organization. They're finding that they need to establish standards, policies, and processes for the usage, development, and management of data. They recognize that creating the right organizational structure and developing the technology infrastructure to support management of their data is critical, too.

As a practice with roots in corporate and information technology (IT) governance, data governance emerged. Data governance can be defined as the processes, policies, standards, organization, and technologies required to manage and ensure the availability, accessibility, quality, consistency, auditability, and security of data in an organization.

# 2 Evolution of Data Management

To understand how data management is evolving to data governance as a practice today, let's review its roots in the broader context of corporate and information technology (IT) governance.

### 2.1 Corporate and IT Governance

An established discipline for many years, corporate governance is defined as 'the set of processes, customs, policies, laws, and institutions affecting the way a corporation is directed, administered or controlled' [1]. One of the practice areas under corporate governance is IT governance, defined as 'the leadership and organizational structures and processes that ensure that the organization's IT sustains and extends the organization's strategies and objectives' [2].

Like corporate governance, IT governance is also a mature discipline. Industry bodies, such as the IT Governance Institute and International Organization for Standardization (ISO), have established detailed IT governance standards, which many companies and government organizations have put into practice.

However, many of these IT governance practices have evolved in a pattern based on common IT organizational structures and deployment architectures, which traditionally have centered on software applications such as enterprise resource planning (ERP) or customer relationship management (CRM) [3]. From mainframes to client-server architectures to Web applications, IT departments have focused on the applications and the user interfaces to those applications, rather than on the data that drives the applications. Typically, data is managed in the context of how it serves a particular application, rather than how it serves the strategic interests of the overall organization.

As a result, IT governance practices are designed primarily around an organization's applications – not its data. In most organizations, there are clearly defined owners, processes, and policies in place to manage enterprise software applications.

# 2.2 The Rise of Consciousness of the Value of Data

An application-centric approach to IT governance no longer serves the best interests of today's rapidly evolving business and IT environment. As shown in Fig. 1, organizations are striving to align their IT systems with their business processes, rather than force-fitting business processes to the constraints of existing applications or systems. As most business processes span multiple applications, IT groups as well as software vendors are busily decomposing applications into components or services that can be mixed and matched in a flexible way to support business processes.

Application-Centric





Moreover, it is becoming increasingly clear that enterprise data – such as customer, employee, or product data – also needs to be shared across multiple applications and business processes, rather than being tied to one specific application. Therefore, to get the most value out of their enterprise data, organizations need to pay greater attention to the data itself and how to manage it as an asset.

Data is one of the most important assets in an organization, even though its value is not represented on the balance sheet. Data is really a manifestation of the business – data represents an organization's customers, employees, and suppliers; its activities and transactions; and its outcomes and results. Data is also the ultimate reusable asset – if it is managed correctly.

For example, data on the profitability of a particular customer can be used in multiple business contexts. This data can drive day-to-day decisions regarding how to prioritize a customer's call in a customer service center. It can feed analyses of product or customer segment profitability. It can drive how the marketing group designs loyalty and retention programs. And it can feed into the finance department's reconciliation of margins and earnings.

To increase the value of data as a business asset, companies and government organizations need to establish standards, policies, and processes for the usage, development, and management of data, to create the right organizational structure, and to develop the supporting technology infrastructure. Thus, data governance has emerged as an important discipline in its own right. The goals of data governance are to:

- Ensure data meets the needs of the business.
- Protect, manage and develop data as a valued enterprise asset.
- Lower the costs of managing data.

A recent survey [4] highlights the increasing number of organizations investing in formal data governance initiatives. Eight percent of the 750 responding organizations have deployed a data governance initiative; 17 percent are designing or implementing one; and 33 percent are considering it.3 The results also make clear that data governance is an emerging practice area – most governance efforts are in the early phases, and many organizations are still working to build the business case for data governance.

# 3 Data Governance

Data governance is a relatively new topic in the business and information technology environment. However, it is a concept that is growing rapidly, as it becomes more evident that sound data management policies can deliver considerable business advantage (streamlined marketing campaigns and enhanced visibility into systems and processes, etc.), and can help avoid serious, sometimes catastrophic problems. It is clear that the impetus for data governance is becoming more business than technology focused.

Poor data oversight and governance is no longer just a technical deficiency companies can choose to live with; today, it can bring about dire legal trouble. For instance, the United States government demands that public companies feature high-quality, transparent data that can be easily tracked. That is, it must be reasonably easily understood, rather than obtuse.

Additionally, when companies submit reports to stock exchange, they must be accurate, or serious economic problems will ensue. To accommodate these mandates, companies must bring data together from often very disparate sources, and make sure it is consistent. In many ways, this task is extremely difficult.

For example, disparate sources are often indeed very disparate; data from different sources usually arrives in significantly dissimilar formats. With multiple sources using multiple values, complexity mounts quickly. Programmers and users, e.g., employ a variety of backend tools and spreadsheets, in addition to other sources.

As such, it is next to impossible for an individual data user to determine with precision how an organization arrived at a particular number, or value. Most of the time, there are simply too many contributors to track, so the user goes with the figure that seems most accurate, or sensible. That inability to clarify and qualify a data element certainly falls short of transparency standards. Unfortunately, this approach is not unique; it is standard operating procedure at a great deal of companies.

These organizations must begin to make that process controlled and understood, and they can begin by establishing ownership of data. In fact, the market recognizes this requirement; many companies are approaching their partners and requesting solutions that will enable them to rein in their data, to make it more strategic and useful. Most of the time, these companies do not use the term "governance," but they certainly understand the problems borne of poor data quality and little data control – ramifications such as significant expenses incurred to remedy data collection and storage processes, inefficiency, poor customer service, frustrated supply chain partners, and as noted earlier, a stock calamity.

# 4 The Most Important Drivers to Data Governance

While companies and government organizations may want to achieve the goals of data governance in theory, they have difficulty justifying the effort unless it has a practical, concrete impact on the business [5]. Putting data governance into real practice requires a compelling business driver. The business driver highlights the importance of data to the business, and motivates both the business and IT to invest in the development of their data.

The most common business drivers of data governance initiatives are:

- Growing revenue.
- Lowering costs.
- Ensuring compliance.

While these are among the most basic goals of any business, it's useful to examine them in more detail and review the integral role data governance can play in achieving these goals.

### 4.1 Growing Revenue

One of the most important goals of almost any business is to grow revenue. And one of the most effective ways to grow revenue is to increase cross-sell/up-sell rates and improve retention among existing customers. To do so, organizations need a broad and deep understanding of their existing customers. They need a "single view of the customer" to be able to provide superior service, and to better target campaigns and offers based on a specific customer's needs [6].

What keeps organizations from having a single view of their customers? Accurate, consistent, and reliable data about their customers and the products and services they buy. Following some common complains that can be heard from users [7]:

- We don't know which products or services to cross-sell to our customers because records of what our customers have purchased and when are spread across six different systems.
- If a customer logs an issue in our support system, it takes a few hours to show up in the customer service representatives' system in the call center. So when talking to our customers, our service representatives don't have really up-to-date customer information, and that can really frustrate both the customer and the representative.
- We lose track of some customers, while others we repeatedly barrage with the same offers. Our customer data is unreliable. Sometimes information is missing, and other times there is conflicting information.
- Our sales, marketing, and service organizations all define 'customer' differently, so we cannot get a global view of our customers across all three functions.
- We're worried about security. Our reputation and our competitive position would suffer if an unauthorized person were to access or change any customer data.

Customer data is often scattered across dozens or even hundreds of different business systems. To resolve these data issues, companies must address the underlying organizational, process, and technical issues related to the data. Data governance provides the framework for addressing complex issues such as improving data quality or developing a single view of the customer at an enterprise level.

### 4.2 Lowering Costs

While the pressure to lower costs is not as intense as it has been in recent years, increasing operational efficiency is still a major priority for most organizations. One of the important ways organizations can reduce costs and increase operational efficiency is to automate business processes. For example, organizations may automate their procurement processes to lower purchasing and administration costs.

While business process automation increases efficiency, problems with enterprise data prevent companies from capitalizing on the full potential of operational efficiency initiatives. Streamlining business processes across multiple financial, human resource, sales, and other business systems requires that the structure and meaning of data be reconciled across those systems—a task that has often been an afterthought in operational efficiency initiatives.

Companies and government organizations of all sizes suffer from data-related problems commonly seen in operational efficiency initiatives, including:

- It is hard to negotiate effectively with our vendors because we have trouble getting at the vendor data from our European subsidiary. It's housed in an arcane legacy mainframe system that we can't access easily.
- We are burdened by stock-outs and excess inventory because we can only run our global inventory reports weekly. They take a whole weekend to process.
- Our suppliers all have different product codes, so we can't get a global view of our supply chain. It's hard to forecast inventory levels or predict delivery schedules.
- Sales sends us forecast data, but no one knows how they arrived at the numbers, so we can't rely on them for planning.
- We need to carefully control which data can be seen by which supplier, so that they don't accidentally see information on one of their competitors.

The need to lower costs is driving projects such as supplier or product master data management, which enable companies to streamline core business processes such as inventory and supply chain management by rationalizing, cleansing, and sharing key master data elements. Data governance plays a critical role in the success of such projects, providing a structure for addressing the organizational and process issues around master data.

### 4.3 Ensuring Compliance

Doing business today requires compliance with a growing number of external regulations, as well as with internal corporate governance policies designed to increase transparency and prevent corporate malfeasance and fraud [8]. To ensure compliance with regulations, as well as with internal policies and controls, companies must streamline the collection of reporting data. For many regulations, they must also document the sources

of the data being reported and certify its accuracy, and implement specific governance policies.

Doing business today requires compliance with a growing number of external regulations, as well as with internal corporate governance policies designed to increase transparency and prevent corporate malfeasance and fraud. To ensure compliance with regulations, as well as with internal policies and controls, companies must streamline the collection of reporting data. For many regulations, they must also document the sources of the data being reported and certify its accuracy, and implement specific governance policies.

Complying with these regulations and policies can burden a company when it comes to how it handles its data. Common concerns include [9]:

- It is hard to pull together the financial data from the dozens of different sources, from mainframe to spreadsheets, so there's always a lot of IT involvement to pull the data, which slows the whole process down.
- Our executives are demanding access to all the information on a daily basis on his computer. We can't do that right now.
- The data we report to the auditors has to be clean and accurate, and right now it's nowhere close.
- Our different business units use different charts of accounts, so it takes days for our analysts to reconcile them.
- A lot of our compliance reporting is done via spreadsheet, which is not going to hold up with the auditors.
- This is sensitive data. We have to carefully control access to it, or we could face huge fines.

Data governance is an essential foundation for ensuring compliance. It establishes the rigorous data standards, policies, and processes that are required by regulations and corporate governance policies, and helps to automate compliance-related tasks, lowering costs. It also helps ensure accountability for and auditability of the data.

# **5** External Influencing Factors

Beyond the business drivers mentioned above, two macro business trends are further escalating the need for data governance: an increase in merger and acquisition (M&A) activity, and a rise in partnering and outsourcing non-core business functions.

### 5.1 Mergers & Acquisitions

As merger & acquisition activity picks up, organizations are faced with the need to rationalize and reconcile the IT environments from merged or acquired entities. Typically these IT environments have very different systems, data models, and business processes. Post-M&A, IT organizations are often pressed to meet very tight timelines for integration. The goal is to accelerate the promised synergies from the merger, both in the form of cost reductions by eliminating redundancies, as well as revenue growth from increased cross-selling.

The process of migrating and consolidating the data after a merger or acquisition is a huge task – one that is often underestimated initially. IT groups must deal with unknown systems, resolve quality issues, and provide detailed documentation on how the information has been merged. And the task involves much more than technical integration.

IT organizations must reconcile different data definitions and models, and processes must be put in place to ensure alignment of the various entities. A data governance framework provides significant value in managing the organizational and technical complexities of merger & acquisition consolidation and accelerating positive business results.

### 5.2 Partnering & Outsourcing

Another broad market trend is the increasing use of partners and outsourcers to manage parts of the value chain. Organizations are focusing on core competencies and handing off non-core functions and processes to partners and outsourcing providers. For example:

- High tech equipment companies rely on contract manufacturers for production.
- Manufacturers turn to specialized transportation and similar companies for logistics and warehouse management.
- Pharmaceutical companies rely on third-party clinical trials management firms.
- Information Technology (IT) departments outsource software, service and application development, as well as network management.
- Human Resources (HR) groups outsource administrative functions such as payroll or benefits management.

As business processes and IT systems shift to outside providers, the data associated with those processes and systems relocates outside the boundaries of the organization. Organizations must ensure that the data is correctly migrated to the outside provider. The data must be complete and accurate, and it has to be restructured to work in the third-party system.

It is important to note that although it has moved to a third party, this data remains a core asset of the organization. Even though it sits outside the firewall, the organization cannot relinquish visibility into and control over that data. A robust data governance framework is critical to managing data that is fragmented across the extended value chain, especially in defining the standards and processes for interaction and collaboration with external partners and outsourcers.

# 6 Data Governance Framework

### 6.1 Major Data Attributes

Earlier in this paper we examined typical data challenges associated with business goals, such as increasing customer cross-sell or ensuring regulatory compliance. By examining the challenges that arise over and over again across organizations, it is clear that enterprise data must have the following six attributes:

- *Accessibility*. Ensuring that all enterprise data can be accessed, regardless of its source or structure.
- *Availability*. Ensuring that data is available to users and applications, when, where and how needed.
- *Quality*. Ensuring the completeness and accuracy of data.
- *Consistency*. Ensuring the meaning of data is consistent and reconciled across systems, processes, and organizations.
- *Auditability*. Ensuring there is an audit trail on the data.
- *Security*. Ensuring secure access to the data.

Data governance manages and develops these key data attributes, enhancing the overall value of the data as an asset to the organization.

### 6.2 Components of Data Governance

To ensure that enterprise data is all of the above accessible, available, of high quality, consistent, auditable, and secure— an effective data governance framework involves four key components:

- *Standards*. A key function of data governance is to establish the standards for data in an enterprise. Companies need to establish data definitions and taxonomies, define master data, develop enterprise data models, and enforce development and technical standards related to data.
- *Policies and Processes*. Establishing and enforcing policies and processes around the creation, development, and management of data is the foundation of an effective data governance practice. Companies need to define data and data-related business rules, control access to and delivery of data, establish ongoing monitoring and

measurement mechanisms, and manage changes to data.

- Organization. Arguably the most important issue that companies must address when launching a data governance initiative is how to design the organizational structure. Companies need to define the roles and responsibilities within the organization that are accountable for the data.
- *Technology*. Hypothetically, companies could embark on a data governance initiative without an underlying technology infrastructure. Indeed, many organizations launch their initial data governance programs using manual tools – spreadsheets or Word documents—to capture data definitions and document processes. However, most of companies quickly realize that this kind of manual approach is severely limited. Indeed, it is nearly impossible to achieve the ultimate goals of data governance using a manual approach.

Technology can help automate and scale the development and enforcement of data governance standards, policies, and processes [10]. Specifically, a data integration technology platform that provides built-in capabilities to access, cleanse, transform, deliver, and monitor data is ideal for data governance. Just as a business process management solution helps to streamline business processes, a data integration platform helps to automate data-related processes [11].

Fig. 2 shows the building blocks of an effective data governance framework. Such a framework takes into account the standards, policies and processes, organizational structure, and technology infrastructure required to make data accessible, available, of high quality, consistent, auditable, and secure across the enterprise.

Data Governance					
Data Accessibility	Data Availability	Data Quality	Data Consistency	Data Security	Data Auditability
Standards		Policies & Processes		Organization	
Data Definitions & Taxonomies	Master & Reference Data	Data Definition	Monitoring & Measurement	Roles & Responsi- bilities	Training & Education
Enterprise Data Model	Technology & Tools Standards	Data Access & Delivery	Data Change Management	Planning & Prioritization	Org. Change Management
Data Integration Infrastructure					

Fig 2 – The Building Blocks of Data Governance

# 7 Organizational and Technology Success Factors

# 7.1 Organizational Issues

Organization is one of the most important aspects of data governance. To achieve the goals of data governance, ownership of the data must be assigned, standards must be defined, and policies must be enforced–all of which can trigger tricky political situations and fierce territory battles [12]. This paper does not attempt to propose an ideal organizational structure to implement or support data governance. However, certain organizational principles are consistent across companies with successful governance programs:

- Clear definition of roles and responsibilities. A fundamental requirement for any type of governance program corporate, IT, or data is clear roles and responsibilities. Companies need to be rigorous about defining roles and assigning specific responsibilities to individuals involved in data governance to enforce accountability [13].
- Business and IT involvement. A common factor among organizations with successful data governance initiatives is that executives and staff from both the line of business and the IT department are involved. In most cases, the business assumes ownership of the data and takes the lead in driving data governance. This is appropriate, since the data ultimately exists to serve the business, and the business is the primary beneficiary of effective data governance. IT then partners with the business to implement the technology aspects of the data governance program. IT and the business collaborate to establish specific business metrics associated with the data governance program and track them over time. Demonstrable business benefits are critical to the success of the data governance program, raising visibility of and increasing support for the program throughout an organization.
- *Executive sponsorship*. As a corollary to the principle above, successful data governance programs are sponsored, ideally, by a senior business executive. Senior executives must recognize the value of a data governance program in managing data as an asset and be able to tie the program to specific, concrete business goals. Many organizations establish a cross-functional data governance executive council or steering committee, which includes both business and IT executives representing various business units and functions.

Integration competency center. Successful data governance initiatives frequently involve the creation of an integration competency center (ICC), or center of excellence, to support the data integration technology that sustains data governance. ICCs are an organizational approach designed to increase agility and reduce implementation costs by promoting reuse, sharing best practices and resources, and establishing common processes and standards for integration. An ICC becomes a shared resource for the entire organization, defining the technical standards and processes around data integration and data governance, and providing a pool of highly skilled technical resources that can support specific project and program implementations.

### 7.2 Technology Issues

A data integration technology infrastructure helps to deliver on the goals of data governance by automating the data integration lifecycle. Data integration is not a linear, one-time occurrence. Data integration is an ongoing, iterative process that constantly seeks to improve upon the key data attributes, such as quality and availability. And as such, when considering data integration technology, organizations need to take a holistic approach.

There are seven key steps in the data integration lifecycle:

- *Access*. All data is accessed, regardless of its source or structure. This includes extracting data out of arcane mainframe systems, as well as relational databases, applications, XML, messages, and even spreadsheets.
- *Discover*. Data sources, particularly poorly documented or unknown sources, are profiled to understand their content and structure, infer patterns and rules implicit in the data, and to flag potential quality issues with the data.
- *Cleanse.* Data is cleansed to ensure its quality, accuracy, and completeness. This may include addressing errors or omissions, enforcing adherence to data standards, validating values, and eliminating duplicate data entries.
- *Integrate*. To maintain a consistent view of data across all systems, data is integrated to bring together fragmented information and transformed to reconcile discrepancies in how different systems define and structure various data elements. For example, the marketing and finance systems may not only have different formats for the 'customer profitability' data element; they

may have completely different business definitions for 'customer profitability' [14].

- *Deliver*. The right data is delivered in the right form, at the right time, to all the applications and users who need it. This can range from delivering a single data element or record to support a real-time business operation, to delivering millions of records to enable trend analysis and enterprise reporting. Delivery also includes ensuring that the data is both highly available and secure in its delivery.
- *Develop and Manage*. A high-productivity toolset enables data stewards, business analysts, architects, and developers to collaborate on the implementation and management of data integration rules and processes, spanning Steps 1 through 5 above. It also ensures the reliability, scalability and performance required to run mission-critical enterprise systems.
- Audit, Monitor, and Report. Data is monitored, and reports on the data are prepared. This includes ongoing measurement of key metrics such as data quality, with an eye toward steady improvement over time [15]. The goal is to track progress on the key data attributes and flag any new issues so that they can be fed back into the data integration lifecycle for resolution and continuing improvement. This step also includes auditingmaintaining a robust audit trail on the data helps to maintain visibility and control, as well as to reduce the cost of future change.

By taking a lifecycle approach to data integration, schematically shown on Fig. 3, the technical capabilities of the data integration platform are brought to bear as part of the ongoing data governance program, with the goal of continual improvement of and accountability for key data metrics.



Fig. 3 – The lifecycle approach to data integration

Organizations need to address the data integration lifecycle holistically, both from a process and from a technology point of view. Often organizations use a collection of different tools to address different stages of the lifecycle in a piecemeal fashion. Different roles as well as different business units may each select different technologies, optimized to meet their own specific, narrow needs.

For example, business analysts may use one tool to profile and understand existing data sources. Data architects may select another tool to cleanse data and improve its quality. IT developers may use yet a different tool to access and deliver data. The net result is a complex web of disparate, inconsistent approaches to dealing with data.

A siloed approach to data integration can quickly derail a data governance program, preventing it from achieving its goals and scaling effectively. The redundant tools and skill sets resulting from these siloed approaches are cumbersome and costly to maintain. But the bigger problem is the inconsistency in the data and the governance processes themselves. Organizations need a unified approach to data across the entire enterprise, and across the data integration lifecycle.

This unified approach enforces data governance policies and processes systematically across the organization and, perhaps more importantly, instills confidence that the data they rely on to make strategic business decisions is consistent and accurate.

## 8 Attributes of Sound Data Governance Programs

Listing the reasons for implementing a rigorous data governance program is relatively simple: determining the critical processes that must take place, and then implementing them throughout the organization, despite potential resistance to change, is far more difficult [16]. However, the chance for success increases significantly when organizations employ executive sponsorship, strategic oversight, tactical oversight and execution, as shown in Fig. 4.

### 8.1 Executive Sponsorship

Like many large-scale, enterprise-wide initiatives that require people to change the way they approach their work, data governance programs need solid and continuous executive support and long-term commitment. Data governance is best facilitated by direct executive edict. At the start of such an initiative, senior managers must make it clear that information is a critical business asset, and data quality is paramount to the organization's success. Compliance with data quality objectives cannot be optional. As soon as one group falls out of alignment, the entire initiative is doomed.



Fig. 4 – Parties supporting Data Governance success

Additionally, if support from the executive suite – specifically, the chief information officer and operations vice presidents – is lacking, the entire organization will follow suit, and failure will result. Ensuring corporate-wide adoption is critical to the success of any broad-ranging data project. Workers employing poor data practices in one function might not be affected by their laxity. However, that indifference will most likely have a significant impact somewhere in the company.

Companies need to become creative to address a problem of data governance in ways that encourage data accuracy and completeness throughout the organization. For example, compensating order-entry clerks for the quality of the data they input, rather than the quantity, is an excellent start. Incentive-based compensation programs are highly effective, especially with transaction-based work. However, they require investments in time and resources, or they will not be effective [17].

Furthermore, widespread and continuous communication from senior management is integral to the success of such programs. Employees at every level of the enterprise need to understand the importance of instilling processes that ensure data quality and accuracy. In addition to benefits for adhering to these policies, there must be consequences for sticking to the status quo – and the workforce must understand those penalties.

In fact, most employees will embrace such changes if they can see the business benefits they will provide. Of course, if they are simply told to adhere to a directive with no apparent value or reasoning, resistance is more likely. Strong encouragement, along with clearly stipulated business reasons from the executive suite, will help the organization accept a new model.

### 8.2 Strategic Oversight

Strategic oversight is the second layer of successful data governance. Because a data governance program can be sweeping, with many people affected, companies should establish a third committee – a steering committee – comprising senior business and technology leaders. These professionals must make sure large-scale data governance programs remain as effective and efficient as possible, over the long term. In addition to keeping data governance programs on track, and making sure they reflect the business goals (even if, over time, they shift), committee members should establish the overall agenda for the data governance program.

That includes several rather basic responsibilities: consider business priorities, i.e., ensure that all data governance efforts have some element of business benefit; understand the costs involved in the efforts to be undertaken; and determine the number and kind of professionals required to carry out the initiatives. Moreover, the steering committee will make funding requests to the executive team, and determine the order in which various initiatives will be embarked upon.

### 8.3 Tactical Oversight

Data governance also has an important tactical oversight aspect, which should be handled by a Project Management Office (PMO). The PMO's responsibilities include directing specific initiatives being executed under the program [18]. The PMO also creates and enhances business processes and technology implementations.

Additionally, its members determine the approaches, techniques and technologies that will optimize the environment. Finally, PMO leaders manage budgets and work streams at a macro level.

### 8.4 Execution

Execution by the development and support teams and operational groups is the fourth layer of a successful data governance program. Development and support teams execute both day-to-day environment support tasks, as well as solution development and improvement assignments for the program. Additionally, they manage to specific service level agreements, and ensure that they are achieved.

Operational groups manage data from the moment it is added into the organization's systems, until it is deleted. They also handle data versioning, archival and modifications – from an automated as well as manual data management perspective. Finally, the operational groups measure and manage data quality from a service-level agreement (SLA) perspective [19].

Zeljko Panian

### 8 Data Governance Maturity

Because of their relative infancy, data governance programs are not widespread, and in fact, are far from routine at most corporations. Even at organizations that have embraced data governance and begun to implement programs, there is usually much to do. In fact, there are four levels of data governance maturity – reactive, aware, managed and optimized – and most companies reside someplace in the first two. The maturity levels are shown in Fig. 5 and detailed in the following paragraphs.



Fig. 5 –Data Governance maturity levels

### 9.1 Stage 1: Reactivity

A company with no formal data governance programs, or one at the very outset of such a program, is in reactive mode – an unenviable position [20]. Such a company has no data stewards or identifiable data owners, and business users are not involved in data discussions. In such an organization, data problems can be dramatic, as such, they are difficult to hide, and often noticed in the marketplace. Other governance-related shortcomings include:

- Data is thought to be strictly the domain of the IT group.
- Data is not profiled for quality, or adherence to business rules.
- Data security polices and life cycle management are poorly defined, if at all.
- A compliance audit would likely fail.
- Business users do not know the best source of data.

Clearly, the reactive state is problematic, and companies should work quickly to remedy as many of these deficiencies as possible.

### 9.2 Stage 2: Awareness

A company that understands the importance of data governance is well on its way to improving its lot from that perspective. At such organizations, data ownership is identified at the departmental or project level, and business users are involved in governance-based on projects. These are steps in the right direction, to be sure.

However, data issues may still create problems and lead to "surprises," but they are less likely to be devastating than those that arise at reactive companies. Additionally, companies in the awareness stage sometimes profile critical data (usually through SQL), as part of a troubleshooting effort, or through project work. They are slightly proactive in their approach.

Life cycle management and security are areas of focus, at least for major data projects [21]. And, depending on the subject matter involved, the organization may pass a compliance audit. Finally, business users know where the best data is sourced, but do not know how it is used throughout the enterprise.

### 9.3 Stage 3: Managed Data Governance

In a managed data governance environment, every new initiative goes through a set of established governance procedures, and new data projects profile data (in accordance with a methodology). As a result, the surprises common at less prepared organizations are an exception. In a managed state, business users play a role in most aspects of governance, ensuring that projects align with the organization's objectives.

These companies are also well on their way to having data that serves as a strategic asset. And, they have likely implemented (and are using) robust data tools, and storing the results they derive. They may also be estimating data transformation projects with better accuracy, i.e., within plus-or-minus 10 to 25 percent. As such, compliance audits are very likely to succeed.

### 9.4 Stage 4: Optimized Data Governance

When a company has reached the preeminent level of data governance maturity – the optimized level – it has established exceptional data practices throughout the organization, and with partners, suppliers, and customers. For instance, data ownership, security and stewardship are identified and understood throughout the enterprise, and data is perceived as a corporate asset.

Additionally, business users are thoroughly involved in governance.

Although the market features only a few of them, companies with optimized data governance programs are also able to tie vendor contracts to data quality. They can also monitor vendor performance easily, and do so frequently. Data security is enhanced, too, and integrated into legal and compliance efforts. Project estimating is exceptional – usually within 5 percent of total costs. Users also know where data is best sourced, and where it is used throughout the company.

In short, these companies are extremely intelligent in the way they work with data. They are constantly certifying it, and ensuring that it remains of high quality. Additionally, they consider data as a critical part of an ongoing strategic program, and feature it in all business decisions.

## **10** Conclusion

Data integration has always been – and still is – one of the most outstanding challenges in business information processing [22]. Evolution of data management theory and practice has passed a long way. Today, the ultimate focus is on data governance.

Companies and government organizations of all sizes are striving to manage data as an enterprise asset, to be shared and reused across multiple software applications and systems, business processes, and users throughout the organization. Data is also the ultimate reusable asset – if it is managed correctly.

To increase the value of data as a business asset, companies and government organizations need to establish standards, policies, and processes for the usage, development, and management of data, to create the right organizational structure, and to develop the supporting technology infrastructure.

Data governance has emerged as an important discipline in its own right. The goals of data governance are to ensure data meets the needs of the business, to protect, manage and develop data as a valued enterprise asset, and to lower the costs of managing data.

The lifecycle approach to data governance is suggested, which should serve as a means for enabling key steps leading to enterprise-wide data integration: accessing, discovering, cleansing, integrating, delivering, developing and managing, and auditing, monitoring and reporting data. By taking a lifecycle approach to data integration, the technical capabilities of the data integration platform are brought to bear as part of the ongoing data governance program, with the goal of continual improvement of and accountability for key data metrics. References:

- [1] Wikipedia, http://www.wikipedia.com, 16. 01. 2007.
- [2] \*\*\*, *Board Briefing on IT Governance*, 2nd Edition, IT Governance Institute, 2003
- [3] Symons, Craig, IT Governance Framework, http://www.forrester.com, 29. 03. 2005.
- [4] Russom, Philip, Taking Data Quality to the Enterprise through Data Governance, *TDWI Report Series*, March 2006
- [5] Wahedra, Anurag, Why Customer Data Integration (CDI) Projects Fail – Part III: Poor Data Governance, http://www.siperian.com, 2006
- [6] Swift, Ronald S. Accelerating Customer Relationships Using CRM and Relationship Technologies, Prentice Hall PTR, Upper Saddle River (NJ), 2001
- [7] Corrigan, David, Achieving advantage with operational customer data integration and management, http://www.ibm.com, 02/2006
- [8] Jones, Don, The Shortcut Guide To Network Compliance and Security, Realtime Publishers, 2005 (http://www.realtimepublishers.com)
- [9] Hurley, Jim, Improving IT Compliance: 2006 IT Compliance Benchmark Report, http://www.symantec.com, 08/2006
- [10] Han Jiawei, Kamber Micheline, Data mining concepts and techniques, Morgan Kaufmann Publishers, 2001
- [11] Shah, Faisal, Data Integration Strategies for Reliable Information Delivery, *DM Review*, 11/2005 (http://www.dmreview.com/editorial/dmreview/)
- [12] Kotter, John, Managing Change: The Power of Leadership, *Balanced Scorecard Report*, Vol. 4, No. 1, January – February 2002, pp. 6-9

- [13] Rigby, Darrell, Management Tools and Techniques: A Survey, *California Management Review*, Volume 43, Number 2, Winter 2001, pp. 43-50
- [14] Solomon, M. R., Rabolt, N. J., Consumer Behavior in Fashion, Prentice Hall, New York (NJ), 2004
- [15] Evans, John, Data Integration Challenges & Benefits for Enterprise Applications Migrations, Consolidations and Upgrades, http://www.ascential.com, 06/2005
- [16] Harrison, Amber, Don't Forget About Change Management, *BI Review Online*, 14. 02. 2008. (http://www.bireview.com
- [17] Kunreuther, Howard, Incentives for Mitigation Investment and More Effective Risk Management: The Need for Public-Private Partnerships, *Proceedings of the World Bank Conference on Issues for a Consultative Group for Global Disaster Reduction*, Paris, June 1-2, 1999, pp 166-172
- [18] Walton, Jane, Creating a Project Management Office (PMO), http://www.serena.com, 01/2007.
- [19] Zhen, Jian, SLA 101: What to look for in a servicelevel agreement, *Compterworld*, 15. 03. 2006. (http://www.computerworld.com)
- [20] Moseley, Marty. Agile Data Governance: The Key to Solving Enterprise Data Quality Problems, DM Review Special Report, 18. 09. 2008. (http://www.dmreview.com)
- [21] Edelstein, Ryk. Data Loss Prevention: Managing the Final Stage of the Data Life Cycle Model". http://www.deadondemand.com, 05/2007
- [22] Friedman, Ted, Beyer, Mark A., Bitterer, Andreas, Magic Quadrant for Data Integration Tools, 2006, <u>http://www.gartner.com</u>, 22. 11. 2006.