White Paper

Kalido Data Governance Maturity Model

September 2010

Winston Chen

Vice President, Strategy and Business Development

Kalido

Introduction

Data management has gone through significant changes in the 20 years that I've been in this business. Data emerged out of the lockboxes of disparate legacy transactional systems, and data management came to be a separate and sophisticated discipline enabled by advanced software and hardware. Gone are the days when most people needed to be convinced that data is a valuable asset. Through three recessions (1990, 2001 and 2008), the data management industry marched forward nearly unscathed. Spending in this sector continues to increase faster than overall IT spending.

Despite high mind and wallet share, most organizations continue to struggle with chronic data problem. The rate at which the problems grow and proliferate is outpacing our current ability to address it. This is why data governance is getting so much attention. Data governance, as a business practice enabled by the right technology, has the potential to create a leap forward in our capability to manage data.

Like any important business capability, data governance requires organization, processes, and technology to be successful. And adoption does not happen overnight. Kalido's data governance maturity model is based on market research with more than 40 companies at varying stages of maturity. Our maturity stages: Application-Centric, Enterprise Repository- Centric, Policy-Centric, and Fully Governed, map to the evolution of how organizations treat data assets. At the same time, is the Kalido Data Governance Maturity Model is prescriptive in that it provides milestones for organization, process and technology – which need to be aligned – to advance to a more mature stage.

	Stage 1	Stage 2	Stage 3	Stage 4
Organization	Nothing	Silo'ed	Formed	Permanent
Process	Nothing	Informal	Defined	Optimized
Technology	Transaction	Data	Data Policy	Policy Driven
	Application-Centric	**Enterprise Repository-Centric**	**Policy-Centric**	**Fully Governed**

Stage 1. Application-Centric

When organizations began mainstream adoption of data processing technology, systems were built to support transactional business processes to make them less labor intensive: taking an order, balancing a general ledger, and other routine business activities. Data was seen as a by-product of operating a business and had little value beyond the transaction and the application that processed it. Data is not treated as a valuable and shared asset, so the need for governance is not recognized.

At this stage, some organizations attempted to govern data through enterprise data modeling. Their success was limited for two reasons: 1. These efforts were driven by IT and without the broad organizational support and authority to enforce compliance; 2. The inflexibility of packaged applications further reduced their effectiveness. So data governance through enterprise data modeling is mostly an academic exercise.

Organization

Authority and stewardship do not exist. The separation of business and IT is clear, and there is little to no collaboration between them. Business views data as the responsibility of IT.

Process

No process exists for data governance.

Technology

Models of data and business processes as well as rules are entirely embedded in applications. Tools for modeling, managing, and quality assuring data do not exist. There's little in the way of repositories that capture enterprise-wide, cross-functional views of data.

Stage 2. Enterprise Repository-Centric

Starting from roughly 1990 (you can debate the exact year), most organizations began to realize that the value of data extends beyond transactions. Decision making increasingly relied on data analysis. Plus, business processes consumed increasingly large amounts of data created in distant parts of an organization for a different purposes. This led to a trend of thinking about data for broad use cases beyond the localized context of a transaction.

We tackled this problem by building large scale repositories, such as data warehouses, that take an enterprise perspective on data. ERP and ERP consolidation — the notion of having a single, integrated set of plumbing run the business — is driven by the same philosophy. More recently, we've accepted the idea that not all data is of equal value, and it is more cost-effective pay extra attention to the data that describe widely-referenced core business entities . This led to the current build out of master data repositories.

While enterprise repositories yield a lot of benefits, they also have the deserved reputation of being very expensive and risky undertakings. Creating a view of data that supports multiple use cases invariably results in conflicts, and ultimately it's up to the business (the consumers of data) to resolve these conflicts. As a result, data governance gradually but surely came to be recognized as critical to success.

However, because data related activities are generally carried out on a system-by-system basis, governance is typically siloed around individual enterprise repositories: data governance for a data warehouse or an ERP system, for example, or data governance for master data management. Also, governance is informal, lacking a distinct organizational structure and clearly defined and executed processes.

Organization

Authority for data exists in IT but wields limited influence on business processes. Informal data "experts" perform some data stewardship, but there's no official recognition of the role and definition of responsibilities. IT understands that business involvement is critical when it comes to data, but collaboration can be inconsistent and generally reliant on a data-savvy champion in the business.

Process

Loosely defined processes exist around enterprise repositories like a data warehouse, master data hub, and large-foot-print operational systems. Data problems are typically dealt with reactively without systematically addressing the root cause. While IT has processes in place for making technical decisions around data, no institutionalized processes exist for making enterprise-wide, business centric decisions for data.

Technology

Data warehouses and master data hubs of varying scope exist, and technology investments, like data quality tools and metadata tools for data are generally made around these systems. Efforts to manage data across multiple systems are generally carried out in a bottom-up way with limited influence.

Stage 3. Policy-Centric

Even though consolidation is the theme of many enterprise data initiatives, in reality, the opposite has occurred. Departmental systems and data repositories proliferated. It's not hard to see why. Data complexity and volume continue to explode. Business has grown more sophisticated in their use of data, which drives new demand that require different ways to combine, manipulate, store, and present information. Merger, acquisitions, and other

strategic business changes lay waste to long term plans in the midst of execution. In other words, enterprise repositories alone have been unable to keep up with business reality.

Forward thinking companies recognized this and began to solve the data problem in a different way. They formed business-led governance organizations to care for data for the enterprise, and created collaborative processes to manage a core set of data deemed critical for the business. More significantly, they took a policy-centric approach to data models, data quality standards, data security and lifecycle management. Rather than envisioning ever-larger and more encompassing repositories, they put processes in place for defining, implementing and enforcing policies for data. It is acceptable for the same type of data to be stored in multiple places as long as they adhere to the same set of policies. Enterprise repositories continue to be important, but they're built on governed platforms integrated with enterprise data policies.

Emerging from this is a shift in mindset: Business takes increasing responsibility for data content, and data is widely recognized as one of the most valuable corporate assets throughout the organization. For IT, a policy-centric approach is liberating. It affords more flexibility in designing systems to serve business needs without giving up consistency and control.

Companies at this stage are generally well-practiced at a subset of data governance in terms of competence. Of the four areas of competence: data model, data quality, data security, and lifecycle management, the first two are the most common.

Organization

A central authority in a form of a cross-functional council is formed and data stewards are explicitly appointed and given clear responsibilities. Business is engaged in a sustained way in managing data. The mindset of data as asset is taking hold.

Process

Processes for policy definition, communication and enforcement are implemented. A clear process for reporting and tracking data issues is established. Rather than having siloed processes, key enterprise data repositories are governed by a single, streamlined set of governance processes.

Technology

A centralized and easily accessible repository of data policies is established to set policies in a top-down fashion, and the process of data governance is supported and orchestrated using automated workflow. Data quality is regularly monitored and measured.

Stage 4. Fully Governed

Successful implementations of policy-centric data governance will produce pervasive and long-lasting improvements in business performance. Over time, the scope of these data governance programs will increase to cover all major areas of competence: model, quality, security and lifecycle. And clearly defined and enforced policies will cover all high-value data assets, the business processes that produce and consume them and systems that store and manipulate them.

More importantly, a strong culture that values data as a strategic asset becomes firmly entrenched and manifested in every aspect of doing business. Like human resource management, a distinct data organization with institutionalized governance processes becomes a permanent business function.

Organization

The organizational structure for data governance becomes institutionalized and viewed as critical to business no different than other business functions like HR and finance. Business takes full ownership for data content and data policy making.

Process

Data governance is a core business process and decisions are made with quantifiable benefit-cost-risk analysis.

Technology

Business policies for data model, data quality, security, lifecycle management are integrated with user interactions with data. The centrally defined policies and rules drive behavior of systems where possible. Data at rest and data in-flight are monitored and issues are addressed proactively before negatively impacting the business.

Recommendation

We suggest that companies evaluate themselves along the 18 criteria in our model and understand where they are. An online assessment tool can be found at the Data Governance Resource Center athttp://www.kalido.com/resource-center-data-governance.htm.

Although the results may not all fall neatly in a single stage, it will provide a baseline. Companies can then look at gaps to fill in order to fully align organization, process and technology and plan for the future by looking beyond for future milestones in order to reach the next stage of maturity. Our model can be tailored to the vocabulary and situation of a specific organization.

Organization, process and technology need to advance in lock-step for data governance to be successful. The journey is not quick or easy, but the benefits can be breathtaking.

Organization	**Authority**	No official authority for data; administrators for applications serve as the closest substitute.	A formal group such as Data Architecture within IT has some control over data but lacks the necessary authority to change business processes.	A council or board with high-level representation from some business functions. The council has the authority to change some business processes.	A cross-organizational council or board with institutionalized, enterprise-wide authority for all key decisions involving data.
	Data Stewardship	No data steward role. Traditional IT is the de facto steward of data.	Informal data experts perform some of the tasks of stewardship, but their roles and responsibilities are not explicitly established.	Formal data steward roles are defined and designated for some key data areas with clearly prescribed day-to-day activities.	Data stewards are clearly designated for all key data areas. Stewards are highly visible focal points for data.
	Business Role	Business has no clear role except to provide initial requirements for application development.	Business fully participates in and sometimes leads projects, but its involvement is project-based rather than permanent.	Business is engaged in a sustained way in managing data and data policies. Some end-to-end process owners take an active role in making data policies.	Business takes full responsibility for data content and for data policy making.
	Collaboration	Rigid boundary exists between business and IT with little collaboration.	Collaboration clearly exists. It is intense during a major initiative but is ad hoc on a day-to-day basis.	Business-IT collaboration on data is institutionalized as a routine activity even in the absence of a major initiative.	Business-IT collaboration related to data is pervasive throughout the enterprise.
	Accountability	Traditional IT is completely accountable for data, but accountability is not aligned with business objectives.	Traditional IT is accountable for data, and accountability is somewhat aligned with business objectives.	Accountability for data and its quality is documented and assigned to the most appropriate individuals, typically not IT. However, there is typically no way to enforce accountability.	Accountability for data is institutionalized with common, measurable performance metrics tied to employee performance.
	Cultural Attitude	Data is a by-product of business activities and not valued until someone needs it.	Intuitive awareness that data is an asset, but the organization lacks a framework to determine the relative value of different types of data.	The concept of data as an asset has emerged; data is valued, and activities are prioritized based on business impact.	Pervasive culture of treating data as a strategic enterprise asset with quantifiable value.
Process	**Policy Management**	No concept of data policies. Rules for data are embedded in application logic and are not accessible.	Loose and informal processes for data governance centered around major systems. The processes tend to degrade over time and are impossible to audit.	Transparent processes for managing cross-system data policies are established. End-to-end process satisfies auditors and regulators.	Data governance, including policy definition, implementation and enforcement is a core business process in its own right.
	Communication	Communication occurs during system deployment and training.	Communication is infrequent and often in response to a crisis. It takes time and determination to discover policies.	New and updated data policies are communicated to the people impacted; they are easily accessible when needed.	Data policies pop up in context when applicable, and users are guided on how data should be created, used and handled.
	Issue Resolution	There is no way to raise data issues.	An official channel for raising data issues exists but is not effective. Most problems are resolved through informal networks.	Issues are recorded, reported and tracked through to resolution by data stewards working in collaboration with business and IT.	Potential issues are identified in real time and remediated collaboratively before they can negatively impact the business.
	Decision Rights	Decisions for data are primarily made by IT.	Decision-making is system specific and unstructured at the enterprise level.	Decision-making is structured and decision rights are clearly defined and communicated.	Decision-making for data is institutionalized and made with full understanding of the quantifiable benefit-cost-risk trade-offs.
	Performance Management	No performance management.	Metrics are system specific and heavily IT operations oriented.	Some operational metrics for data governance program have been established and are tracked. They are tied to business needs.	Key metrics on efficiency and effectiveness are standardized. Actuals and goals are compared for variance.
	Dataflow Transparency	Data authors do not know who will use the data or how the data will be used. Data consumers do not know where the data comes from.	IT has some documentation on dataflows from authors to consumers, but business in general does not.	Dataflows for some core processes are documented and accessible by data authors and consumers so that they're aware of the dependencies.	Full transparency of how key enterprise data assets are produced and consumed. Data's downstream impact is well understood.
Technology	**Data Policies and Rules**	No concept of data policies. Rules for data are embedded in application logic and not accessible.	Policies and rules exist in loosely documented form. They are not managed through a central and easily accessible repository.	Common enterprise repository of data quality policies and rules established, accessible by all stakeholders including business and IT.	Common and pervasive policy layer for data quality, security and lifecycle fully integrated with key systems.
	Process Orchestration	No data governance process exists to be supported.	Informal workflow using office desktop application and general purpose collaboration tools such as SharePoint.	Data governance processes are orchestrated by workflow with automation to guide the day-to-day activities of the extended data organization.	Data governance processes are orchestrated by workflow and integrated with enterprise workflow engine.
	Compliance Monitoring	Data consumers discover data issues during the course of use but don't know who to inform for correction.	IT uses tool-specific features (for example ETL rejection) to detect violations to rules. Data consumers don't have a way to report additional data issues.	Active monitoring is deployed and run regularly on multiple data repositories to assess compliance. Data consumers can easily raise issues.	Data quality and security monitoring against policies in place for all key data elements and run on both stored and in-flight data. Data consumers have in-system ways of alerting data stewards of errors.
	Modeling	Models exist for each application only.	IT produces bottoms-up, inventory-style metadata management that lacks business visibility and control. Top-down models are not actively used.	Unified and business-accessible models for data, business processes and systems strongly influence system development.	Top-down model actively drives the design and behavior of key systems.
	Master Data Management	Master data resides in disparate applications and is unmanaged.	Single or multi-domain MDM (typically for customer or product master data) is implemented but lacks governance.	Multiple MDM platforms work in concert with a central data governance application to implement and execute enterprise data policies for master data.	Master data complies with enterprise data policies and rules at the points of origin.
	Data Quality	Data quality is poor and is not measured.	IT runs data profiling and cleansing in an ad hoc manner and on narrow uses cases at the repository level.	Data quality for key data assets is measured holistically, reported and tracked over time to sustainably improve it.	Data quality metrics are pervasive and presented in context to help consumers use data effectively.
		Application-Centric	**Enterprise Repository-Centric**	**Policy-Centric**	**Fully Governed**

Contact Information

US Tel: +1 781 202 3200 **Eur Tel:** +44 (0)845 224 1236 **Email:** info@kalido.com or visit our website at www.kalido.com