Intelligent Solutions, Inc.

March 2011

Maximizing the Data Warehouse Through Information Integration and Governance

Claudia Imhoff, Ph.D.

Solve your business puzzles with Intelligent Solutions

Table of Contents

Introduction

Companies, large and small, are realizing that their competitive edge, in fact their very survival, depends on being able to understand their customers, markets, suppliers, and operations. This understanding comes from confidence in their data and access to trusted analytics and reports.

But the amount of data (both structured and unstructured) coming into enterprises today supporting these analytics and reports has increased to massive amounts – RFIDs, mobile devices, social media, and other data creating devices have placed a tremendous burden on our data warehouses and data management technologies.

Given the critical need for this data, it is no wonder that Information Governance has been elevated to an important initiative for enterprises. It seems you cannot pick up a magazine or read an online newsletter without it containing some article or reference to information governance.

But what is information governance? The definition I use is:

> Information Governance is the set of people and business processes for controlling and coordinating information throughout an enterprise. It involves program strategy, executive oversight and support, total cost of ownership, prioritization, compliance, and arbitration, as well as appropriate technology to support the creation of trusted data.

Information governance affects all areas of the enterprise including all lines of the business, departments and geographies in which the business operates. It orchestrates the people, processes, and technology that enable an organization to leverage its data as an enterprise asset. The expected results of the governance process are to maintain business innovation and competitive advantages by effective management and utilization of information assets.
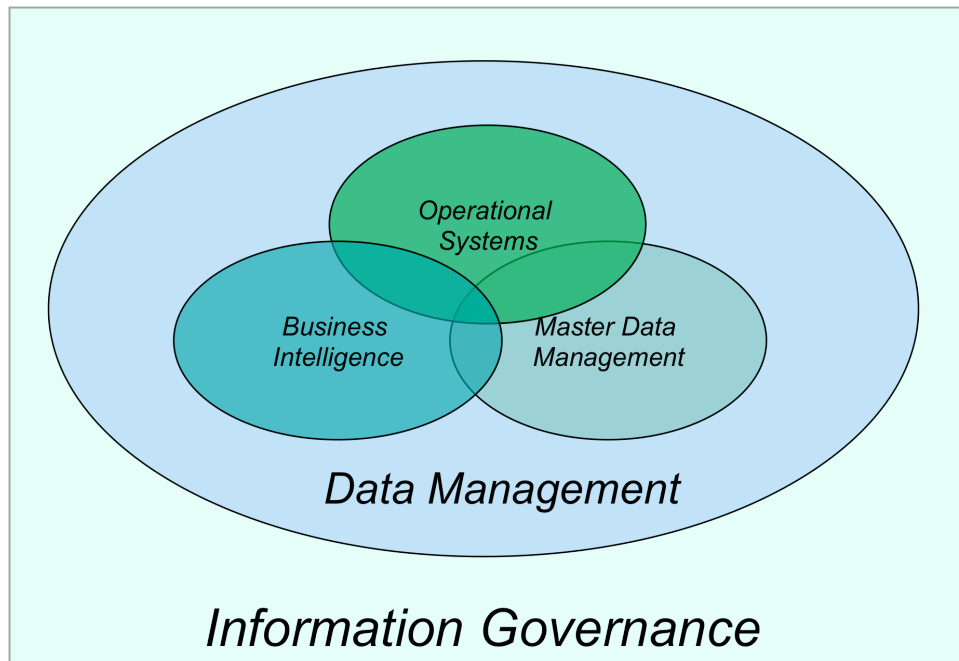
The need for information governance is most commonly uncovered during the design and implementation of data warehouses used for business analytics. Why? Because these initiatives are generally the first to integrate massive amounts of data from various operational sources on a large scale. In fact, much of a data warehouse project's timeline is spent on the arduous tasks of data consolidation and clean up. Usually these projects form informal information governance practices to ensure consistency and repeatability in their

ETL and data quality activities.

However, while certainly useful to the data warehousing and business intelligence efforts, it should be understood that information governance is not limited to this environment. Indeed, it should move to a more formal organization that supports not only business intelligence but also the operational systems and master data management environments. (See Figure 1).

All three environments can benefit from the work products coming from an information governance effort – for example, consistent definitions and formats, sharable data models, standard data integration techniques and technologies.

Figure 1: Information Governance supports the three major IT environments – Operational Systems, Business Intelligence, and Master Data Management

Information Governance Drivers

Drivers for information governance can be divided into four areas – priorities, integration needs, accessibility, and sponsorship,

Priorities

Business priorities often dictate the availability of the critical people and other resources needed for an information

governance effort. This is especially true in constructing a data warehouse. These resources are often the very ones that are already at capacity. They are knowledgeable about how, when and where information is used throughout the enterprise. It is this enterprise view that is another critical factor in the overall success of the information governance initiative and the data warehouse.

If the priority is high enough, then these resources will be freed up to take on the information governance work. If the right people are not available, this is an indication of the perceived priority or lack thereof, and consideration should be given to postponing the effort until the priority reaches the proper criticality.

During the on-going set of projects of the information governance initiative, a limited availability of the right people and resources also impacts the organization's ability to leverage the products from the governance effort for additional value. In addition, this can be detrimental to the overall continuance of the initiative.

Integration Needs

A substantial driver for information governance is the ability to rapidly access integrated data for regulatory and compliance purposes as well as for business intelligence. This data must be trustworthy which could mean several things: A complete data lineage behind it that can be quickly and efficiently uncovered; A comprehensive understanding of data sources and relationships among them; A consistent understanding of key enterprise KPIs between IT as well as Line of Business across all business units and geographies; Standardized, de-duped and cleansed information assets across the enterprise.

Drivers such as these are usually first discovered in data warehousing initiatives and have launched many information governance efforts. It has the advantage of significant executive sponsorship even if the ROI is not immediately apparent.

Accessibility

Many people perceive information as power and some are reluctant to share it. Or they put forth a lot of effort and do not

see a benefit to allowing others access to it. Successful information governance and data warehousing efforts require efficient and secure information sharing. This does not mean that everyone can get to the data. It does mean that those with a legitimate business need are provided with access.

Sponsorship

Having an executive sponsor who is a stakeholder in the outcome of the data warehouse and information governance effort is a must. To get the right support the sponsor should understand the drivers and benefits of the initiative and should have a complete grasp of the effort that will be necessary to implement the full information governance function.

The sponsor can ensure that both the data warehouse and information governance efforts are tied to organizational strategies, which means that the executive sponsor must be high enough in the organization to have a deep understanding of these driving strategies, as well as the ability to effect the outcome of the strategy generation process.

Corporate Information Strategies

Fortunately deployment of a modern data warehouse environment means that much of the technological infrastructure necessary to put in place enterprise-wide information governance initiatives already exists. You have many, if not all, the required data management tools in your technological arsenal. These include data consolidation (ETL), data profiling and quality, as well as data modeling tools. These include the following capabilities.

Data Source Discovery and Consolidation

In most data warehouse implementation cases, clients have a large set of data sources that are in different shapes and formats such as databases, sequential files, spreadsheets, and so on. Understanding how data elements are interrelated across complex and heterogeneous systems is a huge undertaking. The process starts with analyzing data values and data patterns to identify relationships. Then data elements need to be linked into logical units of information (business objects) such as customer,

patient, or payments. These units represent a group of related objects that consist of data from one or more applications, databases or data sources. This step is a time consuming and daunting task that involves a lot of risks and mistakes. An accelerator can increase the time to value in this phase. This task is crucial for the beginning phases of a data warehousing implementation. You will need to leverage technology to speed up this process and automate it. The implementation of the warehouse then starts with these business objects as a foundation for data modeling.

When additional data sources are being integrated into existing data warehouses, you will need to use detection and discovery technologies to understand the structure and semantics of the new and unfamiliar sources as well as to determine how they can be incorporated into the data warehouse.

KPI and Metric Definitions

Before implementing a data warehouse model, Business and IT personnel need to collaborate and agree on essential business metrics and key performance indicators (KPIs). Often a glossary of critical business concepts, KPIs, and metrics provides a central and authoritative source of information. This is an important part of the process to carry forward as long as the data warehouse exists. Misinterpretation of business needs can result in a data warehouse that is not based on the needs of the business. Technical teams responsible for implementing and delivering such a data warehouse need to have that information clearly defined and stated. For example, not knowing that the business has changed the definition of "Gold Value Customer" from a client that owns 3 or more products (e.g., Loans, Checking Account, and Credit Cards), to a customer that is worth more than $100K in total assets, can impact the underlying implementation of the data warehouse.

Having a Glossary that defines terms precisely and accurately with assigned data stewards takes the guesswork out and eliminates risks and redundancies that may occur from miscommunications or ambiguity. The recommendation is to bring all the stakeholders of each data warehouse project around the same table and facilitate their discussions and agreements on building blocks (KPIs and critical business elements). These agreements can be stored using appropriate

technology in a central place with easy web-based access. There is no question whether somebody has the current version of the information. As a result, these business users become invested and involved in the implementation of the project, which translates into a higher adoption of the warehouse once the implementation is finished.

Data Cleansing and Standardization

Trust in the data presented by the data warehouse and established by the information governance function is essential to the organization since the data warehouse serves as the analytical knowledge base for many critical business decisions. Accuracy, consistency and completeness of information in the warehouse play a key role in giving the users full confidence in the data. Visibility to trusted upstream sources of information, and downstream applications and BI reports that consume that information serves to enhance the veracity of the data.

Again foundational technology not only supports the information governance initiative by boosting the quality of the data in the data warehouse, but it establishes benchmarks for monitoring and assessing this quality overtime. These technologies provide visibility to the entire information flow from the sources of information all the way down to the information loaded into the data warehouse and ultimately to the analytics and BI reports. This end-to-end lineage view allows the business users to identify bad sources of data and their impact to downstream applications. When changes are required to reduce the population of bad data, impact analysis capabilities reflect and highlight the risk areas so changes and modifications to the data do not interrupt systems and production environments.

As a prelude to implementing information governance and constructing the data warehouse, it is important to set up a corporate data strategy. You may have started this process during the implementation of your data warehouse but now it is time to formalize the function. The strategy consists of guiding principles for governance and standard data management processes.

Guiding Principles

Guiding principles are not cast in stone – they are living and breathing guidelines that change as the enterprise changes.

Establishing a baseline allows the enterprise to assess the current state of each principle against the desired state for it. Also for each principle, there should be corresponding metrics to determine if the desired state is being reached.

A starter set of guiding principles is listed here.

- Architecture – Structuring data warehouse stores and designing data movement between those stores for accessibility and scalability

- Metadata – Capturing and providing access for data lineage, definitions, quality metrics, transformations, etc.

- Information access and delivery – Providing Information to business units in a timely and business-user friendly method

- Information management – Developing and maintaining enterprise data models for use in data warehouse construction

- Security, privacy, compliance – Securing data, complying with reporting and regulatory requirements, restricting data access, and monitoring information use throughout the data warehouse environment

- Stewardship – Establishing formal data steward roles

- Data Quality – Establishing and monitoring quality metrics and implementing continuous improvement for both the data warehouse and upstream operational systems

- Data as an asset – Recognizing that information is a strategic asset with policies and controls

The information governance group should develop a customized set of principles that fit their culture, information requirements and technological capabilities. If the data warehouse environment is not very mature, these principles and resulting work products can assist the data warehouse developers in their overall implementations. In any case, it should be noted that guiding principles will change over time as the enterprise's situation changes.

Data Management Processes

The next piece of the strategy puzzle is to define the standard data management processes either in practice or desired to be

in practice. Again, a starter set is offered but the information governance function must customize these to fit their enterprise in general and their data warehouse projects specifically.

- Quality Improvement – Establish quality targets, investigate problems, root causes, suggest remediation, monitor, implement solutions back to the operational systems

- Model Management – Develop subject, business and technology models, synchronize with physical BI data stores, incrementally add/change models

- Data Definition – Establish enterprise definitions, identify line of business definitions for preservation, develop policies, identify process changes required

- Prioritization – Submit projects for quality improvement, determine impacts and business value, add to priority queue, approve, decline, or send back for re-work

- Metadata Management – Develop a metadata repository, determine new sources and develop meta model and repository extensions, determine permissions and monitor usage, make assessable to all data warehouse business and IT personnel

- Data Criticality – Determine grading system for data elements, impacts and handling for graded data and methods to change ratings

- Access – Identify data owners, determine access, process access requests, submit for arbitration

Again if you have a well thought-out data warehousing environment, many of these processes are already in place. If not, they can greatly help speed up and ensure the success of future data warehouse projects. We are fortunate today to have technological support for these activities as well.

Getting Started

The following Practical Tips for starting your information governance function have been consolidated from work with a number of companies. They represent the best practices for a modern information governance organization.

- **Tip 1: Determine the personality of your organization.** Each company has distinct characteristics and traits. Some are set up similar to the military where the top executives make all the decisions and the rest of the company personnel simply follow. Others are more matrix-oriented where decisions are made by consensus. Other factors to consider for information governance and data warehousing include whether your company is geographically dispersed, if it has independent lines of business with minimal overlap or interactions, and whether work is centralized or decentralized. Both data warehousing and information governance must have cooperation across the enterprise, which means that how the enterprise makes decisions is a critical component in the success or failure of these initiatives.

- **Tip 2: Develop a repeatable process.** This starts with a definition of the business problem, then obtaining sponsorship. For example, as part of IBM's Information Governance offering, they have outlined a series of process steps that support the governance goals, assess the maturity of the function and the skills needed, generate a starter set of metrics, and establish the roadmap or blueprint for the overall activities. For each project within the function, this process is repeated over and over. It is also important to embed a continuous improvement mechanism in the overall process to ensure that lessons learned are incorporated in the next process.

- **Tip 3: Assess the current state of each guiding principle.** Compare where you are today to the ideal that you documented. This baseline will serve as starting place for the metrics you develop to track progress toward the ideal. The metrics developed are generally some form of key performance indicators or governance scorecard and should be aligned with the overall corporate strategies. These metrics can be mapped to a maturity model to provide a tracking mechanism for progress.

- **Tip 4: Select appropriate team members.** Information governance takes time and patience. In many cases, the people populating this function have a difficult row to hoe. They must be people who support adherence to the enterprise view of data and can resolve business integration issues. They have to be able to balance

individual needs with those of the enterprise. Important traits for information governance personnel include an unswerving respect for the enterprise view of data, a commitment to managing data as an asset, and a personality that can handle delicate negotiations with difficult factions in the enterprise. They must have solid conciliation and facilitation skills to navigate the tricky definitional and business rule disparities.

- **Tip 5: Communicate the principles and processes to the rest of the enterprise**. This helps to achieve the desired results. Specify the types of communications (newsletter, emails, status meetings, web posting) to be used and their purposes. Identify the recipients and coordinate the delivery of communications centrally. All people affected by governance processes must understand the policies and the reasons for them. Some of the most effective communication comes when it is performed by non-team members such as executives talking to each other or governance sponsors talking to other business people.

- **Tip 6: Identify the required tools and technologies to support the data warehouse and information governance functions.** As mentioned earlier, if a mature data warehouse environment exists, many of the tools and technologies needed for governance are already in place. These include data modeling/design, data profiling, data quality, data quality monitoring, and data tracking/auditing capabilities. Look for vendors who have a long history supporting data governance along with the fully integrated platforms to support the function.

- **Tip 7: Start with a small project having a narrow scope.** Because information governance is a long-term program, not a single project, the first project usually has to take on a lot of administrative activities associated with setting up a long-term initiative. Many activities may not be directly involved with the scope of data for the first project; they will have to be performed nonetheless to support all future projects. These include setting up a common vocabulary, gaining knowledge of existing systems and information, creating a data model if none exists, determining security policies and developing administrative procedures for the creation, update, and deletion of data, business rules, and other associated entities.

- **Tip 8: Monitor or audit your progress.** Without auditing, it becomes almost impossible to determine if any progress has been made. Furthermore, it is nearly impossible to learn from missteps if you can't determine what caused them. Audits should not be performed once a year. Audits should be performed continuously and should examine not only what happened but also why it happened. Each part of the information governance function should be audited to ensure that policies are working as planned, that responsibilities do not overlap or create gaps in coverage, and that there is a proper separation of duties.

- **Tip 9: Look for "hardened" industry specific packages.** These can significantly boost productivity, reduce time to implementation, and improve the overall likelihood of success. Packages should be based on a history of involvement in the data and information governance community. They should contain (at a minimum): preconfigured, fully scalable platforms supporting all technology needed for this initiative, demonstrated industry expertise, a standards-based set of models and templates, embedded data quality processes, and a rich set of starter BI reports analyzing and measuring the metrics for information governance. IBM's InfoSphere Warehouse Packs are examples of such packages.

Summary

Information governance is unquestionably the key to protecting one of the enterprise's most valuable assets: its data. A fully functioning information governance function ensures high quality information for *all* uses – operational processing, master data management, and of course, data warehousing and business analytics. It reduces the time and cost in terms of the storage, management and delivery of data to these three technological environments. Further it ensures that this asset is fully protected in terms of security, privacy and compliance with regulatory bodies.

By following the practical tips, purchasing the best technological platform, and seeking the help from industry experts in this area, you will develop a world-class information governance function that can grow and support your company for many years to come.