WHITE PAPER

INTELLIGENT BUSINESS STRATEGIES

Succeeding with Information Governance Using IBM Technologies

By Mike Ferguson
Intelligent Business Strategies
December 2010

Prepared for:

Table of Contents

WHAT IS INFORMATION GOVERNANCE?

Information consists of structure, semi-structured and unstructured data

Every business, large or small, has to process business transactions and undertake reporting and analysis to support decision making. These every-day business activities are dependent on one thing, data. In many cases the data needed is structured data that resides in databases or files. However semi-structured (e.g. XML), and unstructured data may also be required. This combination of structured, semi-structured and unstructured data gives rise to a broader term referred to as *information*.

Business confidence stems from the use of trusted, high quality information

For information to be trusted it must also be commonly understood

Whatever, the information needed, the business confidence in using it is heavily dependent on that information being trusted. It goes without saying that for information to be trusted its quality needs to be high. However trust is not just about the quality of data. If information is used across multiple departments (i.e. shared), then it also needs to be commonly understood wherever it goes. Master data is a classic example of this because it is used by many applications. In addition, sensitive information must also be protected and access to information must be secure. To make all this possible, information must be *governed*.

Information governance (IG) describes the overall management and control of information and can be defined as:

Information governance is about using people, processes, policies and technologies to manage and protect information

"The people, processes, policies and technologies used to formally manage and protect structured and unstructured data assets to guarantee commonly understood, correct, complete, trusted, secure and findable information throughout the enterprise".

For information governance to work, accountability and approval controls need to be introduced into key data and content management disciplines to help manage information. These disciplines include:

Accountability and control needs to be introduced into all data management and content management diciplines

- Data naming and data definitions
- Data modelling
- Enterprise data quality
- Enterprise data integration
- Master data management
- Enterprise metadata management
- Information lifecycle management
- Data privacy
- Data access security
- Taxonomy design
- Content authoring
- Enterprise content management records management

It is also about standardising, integrating, protecting and retaining information to avoid risks and to remain compliant with regulations and legislation.

This paper primarily focuses on the governance of structured data and therefore does not address areas the disciplines associated with taxonomy design, content authoring or enterprise content management shown above.

WHY INFORMATION GOVERNANCE?

Data distribution, redundancy and inconsistency results in increased complexity

The upsuge in appliances is causing 'islands' of data to emerge

The arrival of cloud computing means that some data may now reside outside the enterprise

There are many reasons why information governance is needed. A key reason is the current level of data complexity in many organizations. Data is often scattered across multiple operational and analytical systems that are often not well integrated. For example, master data in many organizations is created and maintained across multiple systems each with their own identifiers, data names and overlapping subsets of the same data. Advances in technology are also adding to the problem. More and more, 'appliances' are being bought to handle specific workloads and causing 'islands' of data to be created. Also, the arrival of cloud computing is encouraging companies to create some applications off premise instead of on-premise resulting in some data being created and maintained outside of the enterprise. Unstructured information is also heavily fractured with valuable business content residing on many hundreds of file servers often not even findable. It is not surprising therefore that the task of governing and managing data is becoming increasingly complex. This is set to continue as data becomes more distributed.

There is a need to increase confidence in information to support decision making

Compliance and risk management are also high in the agenda

A recent 2010 information governance survey[1] of 407 respondents from across the world indicated that the top two concerns were the need to increase confidence in information as the basis of decision making (59% of respondents) and to improve data quality/trusted information (57%). The same survey found two additional areas of concern were fulfilling compliance and regulatory requirements (50%) and lowering business risk (48%). Note that neither cost reduction nor revenue growth are among these top concerns suggesting that profitability is not a key business driver.

65% of organisations are now implementing or planning to implement an information governance project

In terms of implementation status, two thirds of the 339 respondents who answered the implementation status question in the same survey[2] indicated that they are either currently implementing or planning to implement an information governance project within the next 18 months (See Figure 1).

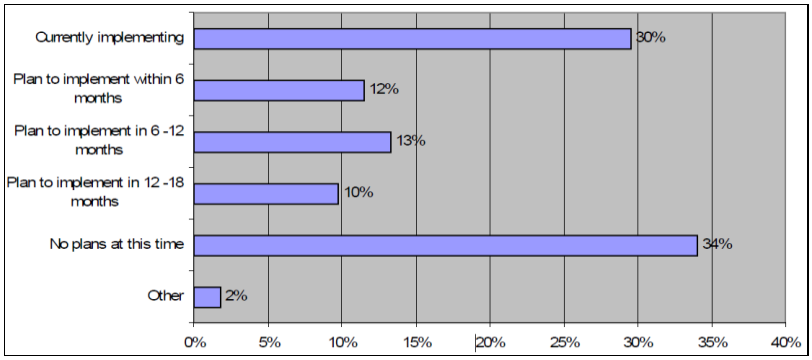

Figure 1

[1],[2] Information Governance as a Holistic Approach to Managing Information – BeyeNetwork Custom Research Report by Judith R Davis 2010

INFORMATION GOVERNANCE SCOPE

Scope should be restricted at first and then incrementally broadened to gradually bring core data under control

As far as scope is concerned, information governance should ultimately span all on-premise, hosted SaaS and cloud based systems as well as B2B exchanges.. However scope can be restricted in a number of ways and an incremental approach taken to gradually broaden the scope to get information under control. For example it could be restricted to:

- Governance of structured reference data only (e.g. code sets)
- Governance of structured master data
- Governance of all structured data in use within a key BI system
- Governance of all structured data across all systems in use within a core operational business process
- Governance of all information in use within a specific business process
- Governance of all information in use within a business department
- Governance of unstructured information e.g. documents, web content, rich media content, email etc.

Information governance should align with business priorities

The scope of an information governance project should be business focussed so that it aligns with strategic business priorities. As scope is broadened this alignment must continue to be maintained so that improvements in the quality, understanding, lifecycle management, protection and security of information all continue to help contribute to high priority business goals. Given the high importance placed on decision making (as shown in the previous section), this may well focus an information governance project on structured data that is used to build BI systems. In the context of business operations then clearly it is the business process that must be considered e.g. an order-to-cash process. In this context information governance focuses on the master data, business intelligence and unstructured content needed in each operational process activity so as to maximise efficiency and straight though processing while reducing operational cost

Structured data governance is often linked to business intelligence

Structured data governance should encompass master data, transaction data and business metrics

In the case of structured data governance, it helps to break this down further into specific types of structured data that need to be governed. For example, this could include:

- Master data, e.g. customer, product, supplier, location
- Transaction data, e.g. orders, shipments, deliveries, purchases, etc.
- Metrics that measure business performance

In this way a systematic and repeatable approach to information governance can be applied to each specific type of data to bring it under control. In the context of master data this allows governance to focus on specific entities, e.g.

- Governance of customer data
- Governance of product data
- Governance of supplier data
- Governance of employee data
- Governance of location data

REQUIREMENTS AND IMPLEMENTATION

Nine key requirements for implementing information governance

Key requirements in any information governance implementation include the need to:

1. Create a vision and strategy for information governance
2. Create the right organisational structure to govern information
3. Define the data to be governed
4. Create the right processes to govern information
5. Define policies and policy scope to govern specific data items
6. Follow an implementation methodology to get your data under control
7. Use technology in each step of the methodology to help implement the policies and processes to govern the data
8. Use statistics and analysis in a scorecard to track progress, monitor issues and monitor continuous improvement
9. Communicate and collaborate with others to gain mass participation (e.g. in common data definitions), maintain interest and commitment as well as for training and awareness

Let's take a look at these in more detail.

CREATING AN INFORMATION GOVERNANCE STRATEGY

A information governance strategy should define objectives, targets, success metrics and initiatives to achieve trusted data

It should also make people accountable

In order to implement information governance, organizations need a strategy. An information governance strategy should set out a vision for information aligned with corporate business strategy objectives and priorities to make sure that implementation focuses on areas that help contribute to achieving high priority business goals. Just like a business strategy, an information governance strategy should define objectives, targets, key performance indicators, ownership, accountability and initiatives to try to reach the targets set. These initiatives are associated with specific data and getting that data under control. For example, there may be an information governance initiative associated with getting customer master data under control. Alternatively, in an insurance company, there may be an initiative underway to get claims transaction data under control in order to help gain a better understanding of risk management and/or meet Solvency II compliance regulations.

In order to measure the success of a strategy it is important to be able to visualise how well you are doing. Information governance scorecards and dashboards are therefore very much a part of implementation. KPIs (success metrics) help you monitor how well you are doing against the objectives and targets set in the information governance strategy to determine if information governance policies and processes are working.

Another important element in an information governance strategy is being able to assess where you are in terms of your ability to govern information. A best practice is therefore to benchmark your organisation against an information governance maturity model in order to plot where your

company currently is on the model versus where you need to be so that a roadmap can be determined. As an example, in the content of master data, a customer master data governance maturity model might look like that shown in Figure 2.

Benchmarking your company against a maturity model helps you plot where you are and dertermine a roadmap going forward

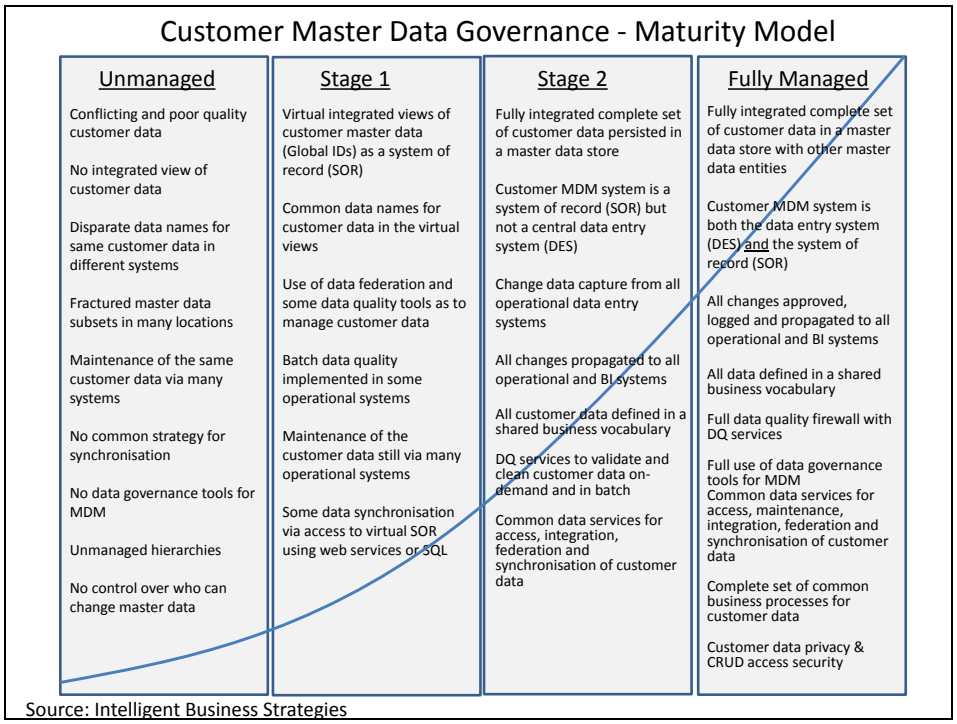

Figure 2

In a broader context, the IBM Information Governance Council also has an information governance maturity model. This has five levels of maturity and three main entry points: information quality management, information lifecycle management and information protection management. More information can be found at www.infogovcommunity.com.

Once a strategy has been defined, people, processes and technology are then needed to implement that strategy. These are discussed below.

CREATING THE RIGHT ORGANISATIONAL STRUCTURE

Organisational structure changes are needed to introduce accountability and control around data

From an organisational standpoint, implementing information governance requires that business and IT appointments are established to introduce accountability and control around data. From a business perspective there is a need for

- A key business sponsor
- An information governance council
- Business data stewards

C-level sponsorship is needed

A business sponsor is a c-level executive in the enterprise who is ultimately responsible for guaranteeing that information is correct. One or more executives could be responsible for this and so it is possible to have more than one executive sponsoring an information governance program. For example the Chief Executive Officer, Chief Risk Officer and Head of Compliance could all be sponsors.

A business led information governance council allows the business to control the approval of data being shared across the enterprise

The purpose of the information governance council is to add controls into the process of defining data for use within core business operations, managerial business processes and in meeting compliance regulations. This is particularly important for data that is shared across the business and sent to regulatory and legislative bodies. This includes being responsible for the approval of new data items used in the business, requests to de-commission data items and changes to existing data item definitions. Change requests should flow through this body for approval before they are submitted to IT. This council is also a business 'gatekeeper' for the enterprise shared business vocabulary (SBV). The SBV is an enterprise wide set of common data names and data definitions for data used in the business and is typically held in a business glossary. A business glossary is a tool that is primarily used by business professionals but is also accessible to IT staff. The information governance council is also responsible for setting and approving policies around information. It is typically made up of business executives that represent core parts of the organisation.

In addition, companies need to assign business data stewards to monitor and govern data used in a specific part of the organisation (e.g. European sales) or associated with a specific business entity e.g. Customer. A data steward is a role. It is not a person. This role can be assigned to a person or to a user group. In the case of a user group, there can be multiple members in a user group, in which case they would all be given authorities and decision rights pertinent to the scope of responsibility assigned to that data steward role. The scope and responsibility of a data steward may be restricted to data for a business area or a specific data entity. A data steward is responsible for monitoring the health of data in their area of responsibility. This includes participating in any data cleansing processes, and taking actions if the quality of data deteriorates. They are also responsible for issuing change requests to the information governance council to approve new data items, changes to existing data item definitions and to request that data items be de-commissioned. In addition the data steward also has access control over users that need to make use of data items within their area of responsibility.

Business data stewards are responsible for monitoring the health of data and managing policies around that data

From an IT perspective, a number of people may be involved in implementing an information governance initiative. These include:

- Data architects
- Data modellers
- Data quality developers
- Data integration developers
- Enterprise content management developers

Chief Data Architect is also a key role within IT

The Chief Data Architect also needs a mandate to get things done in IT

It is likely that a Chief Data Architect will carry responsibility and be accountable for implementing information governance in IT. This role must have a mandate to bring IT into line on all data related issues. In addition, data modellers, data quality developers, data integration developers need to come together to work as a team so that they can set standards and provide a development framework to facilitate re-use of data quality and data integration services across all IT projects.

DEFINING THE DATA THAT NEEDS TO BE GOVERNED

Data that needs to be governed must be formally defined

For any information governance program to be successful, the data that needs to be governed must be formally defined. This is because all business users need a place to go where they can see what data exists in the enterprise, what it means, who owns it and what policies have been applied to it. In addition, users also need to know where that data is actually stored (e.g. application databases, XML files, spread sheets etc.) in the enterprise and where it is used in order to govern it. It is also likely that subsets of the same data will be used in multiple underlying systems. This is due to the fact that data flows across organisational and systems boundaries as enterprise business processes (e.g. order to cash) execute. Therefore, the same data items could exist in multiple underlying systems under the same or different data names with potentially inconsistent integrity constraints as well as differences in other policies. The information governance challenge is to eliminate these inconsistencies and uphold common data names, common integrity constraints and common policies for data no matter where it is used. It follows therefore that data needs to be defined in an application-independent way and then common policies introduced that can be applied to govern subsets of that data in one or more underlying systems across the enterprise. To make things manageable, the scope of specific information governance policies may be restricted to specific organisational areas or systems. Nevertheless, everything needs to be anchored on an SBV of commonly defined data names and definitions.

You should also know where the data is located if it is to be governed successfully

Common data names and data definitions are a critical success factor

In terms of definitions for data, each data item should have:

Every data item needs to be defined

- A data name (term) and data definition
- A description
- Integrity constraints (policies)
- Approved synonyms
- Related terms
- Languages it must be rendered in
- A sensitivity classification, e.g. low, medium, high
- An assigned data steward
- Data quality policies

Policies need to be defined so that data quality, security and privacy can be managed

- Data privacy policies
- Data security policies
- Data lifecycle policies around who is authorized to create, read, update and delete
- A history of who changed names, definitions, integrity constraints and policies, when that happened and what the before and after versions were
- Where the data item is used, e.g. tables, reports

It should be possible to categorise data into business entities and other types of category

In addition, it should be possible to categorise each data item so that different categories exist for data. An obvious example is to place all data items in a business entity (e.g. asset, customer, product, order) in the same category. In this way just looking at the customer category would allow users to see all customer data items. Another example is to place all data items used by an organisational area (e.g. European sales) or business function (e.g. Finance) in the same category. It is also possible to categorise

data items by business process or by process task so that it becomes very easy to see what data is 'in play' during a particular process.

INFORMATION GOVERNANCE POLICIES AND PROCESSES

With respect to information governance *policies* (rules), the first thing to recognise is that there are several different types of policies that need to be defined to govern data. Information governance is not just about data quality. These types of policies include:

Several types of policies need to be defined to govern data

- Data integrity constraints
- Data validation policies
- Data cleansing policies
- Data integration policies
- Data provisioning policies, e.g. Language rendering, formats etc.
- Data access control policies for information protection
- Data lifecycle management policies (CRUD authorisation)
- Data privacy policies for information protection
- Data retention and expiration policies
- Data archive and backup policies

Secondly, these policies can be defined at the individual data item level, at the data category level or both. If they are not applied to a category level (e.g. customer) then all policies associated with data items in that category will be enforced.

Policy scope can be restricted to limit governance capability

Thirdly, policies created by authorised data stewards may also be restricted by *scoping limitations* to limit their governance capability e.g. to allow certain data policies to be set for data in a specific application used in a specific business process by people in a specific organisational area. However if data spans organisational units and systems then enterprise wide common policies may also be enforced.

As far as governance around policies is concerned, it is important that only authorised data stewards and IT professionals should be able to create and maintain these. When they do so, the technology should log details of whoever changes policies and when those changes are made so that all changes can be reported and audited.

Information governance *processes* need to span all data management disciplines including

Information governance processes need to be integrated into all data management disciplines to introduce control and accountability

- Data naming and data definitions
- Data modelling,
- Data discovery
- Enterprise data quality
- Enterprise data integration
- Master data management
- Enterprise metadata management
- Hierarchy and metrics management
- Enterprise content management
- Data security and privacy
- Data backup and archive

The key is that information governance becomes integrated into each discipline so that every aspect of data management is controlled, tracked

and auditable. This also introduces accountability in IT data management practices.

The introduction of new data items, changes to data item definitions and data de-commissioning need to be formally controlled

For example, with respect to data naming and data definitions, this is very much business led discipline. Business users decide what data items the business needs, data names and valid synonyms, their integrity rules and any data item descriptions that say what each item means. However an information governance process needs to be wrapped around this discipline so that there is a formal process for introducing new data items into the company, a formal process for decommissioning data items and a formal process for changing data item names, definitions and integrity constraints. In all of these cases, no matter which business area issues a data request, all *candidate* requests should flow to the governance council for approval as part of an information governance process. If approved, then the appropriate change is *accepted* and entered into the enterprise SBV using a business glossary tool. Once all systems that need this data have been changed, then it becomes *standard* or *decommissioned* across the enterprise, depending upon the type of data request. The reason for integrating information governance here is to make sure that there are formal controls around change. For example:

- Only authorised business users can issue requests for new data items
- Only authorised business users can issue requests to decommission data items
- Only authorised business users can issue requests to change data item definitions
- All changes flow to an information governance council for approval
- The approval process effectively means that each data item can be versioned such that each version represents a specific state, i.e. Candidate, Accepted, Standard, Decommissioned
- A formal record is kept on who requested the change, who approved it, when it was approved, what uses it, etc. to make it auditable

The same applies to data quality rules, data transformation rules, data privacy rules or indeed any policy associated with data.

A Best Practice Methodology for Information Governance

For information governance implementation to be successful, people, processes and technology need to work together in a continuous improvement methodology, that gradually brings data under control. An example of such a methodology is shown in Figure 3.

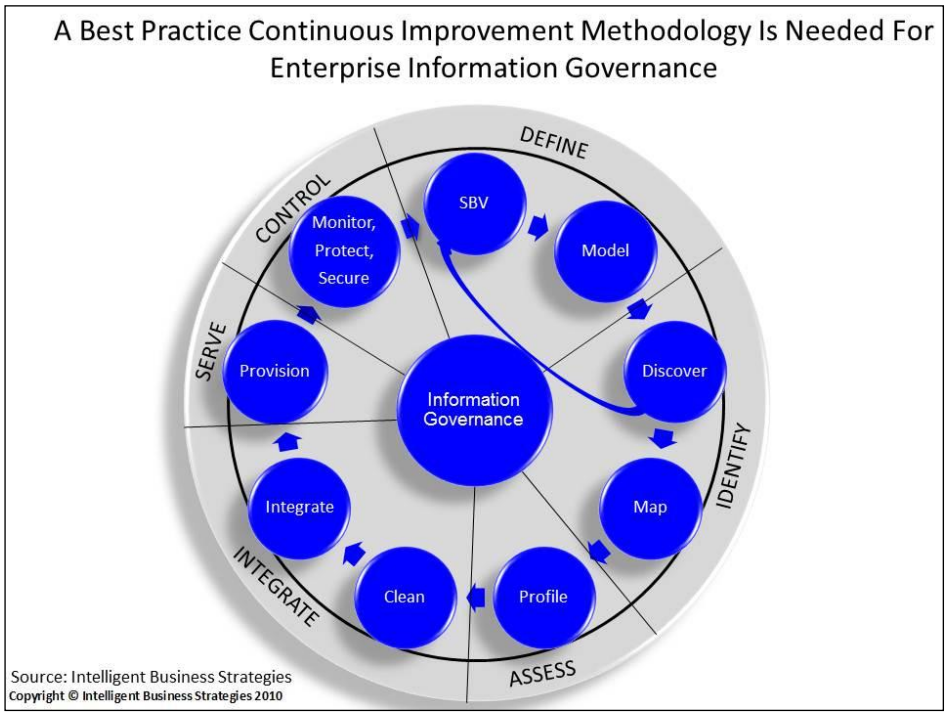

Figure 3

A continuous improvement methodology is also needed to get data under control

Common data names and definitions for each data item are the starting point in an information governance implementation

This implementation methodology starts with defining enterprise wide standard data names and data definitions (i.e. SBV) for core master and transaction data. For example, for supplier data, standard data names and definitions can be defined for all attributes that describe supplier. Similarly for orders data, standard data names and definitions can be defined for all attributes that describe an order. These are defined in a *business glossary* and will follow an information governance process to be approved by the governance council. Once this is done, it becomes possible to create logical data *models* using standard data definitions in the business glossary. For example for customer data, the model may contain either the entire customer entity or some subset of it. The key point here is that the data modeller has to use the business data names and definitions in the business glossary when building a data model.

Business defined common data names and definitions should also be used in data models

Data discovery is used to identify where data entities currently reside

Once this has been done, users can make use of data *discovery* technology to discover what data and data relationships currently exist in the enterprise, and what systems the data is currently located in. This discovery exercise could include discovery across on-premise applications and off premise cloud based applications. It may also be the case that the data discovery activity discovers more customer data items than are currently defined in the SBV. In this case users may wish to go back and extend the SBV, then reflect the additional data items in the logical data model and repeat data discovery to discover the complete set of customer data needed across all systems. At this point users can then *map* the discovered data sources to each other and to the target model, discover the transformations needed and profile the discovered source data to identify data quality problems that exist. From here it becomes possible to generate data integration workflows to *clean-up, transform and integrate* the discovered data to get that data into a state that is 'fit for business use'. It can then be provisioned either into a database or to applications and tools in response to an on-demand request. Once this has been done, it also becomes possible to

Mapping discovered data to common data definitions helps users understand what data means

Profile discovered data to assess data quality

Generate data integration jobs to clean-up and transform data to make it fit for business use

Data privacy is also part of information governance

monitor that data to make sure quality does not deteriorate. Data also needs to be *protected* to maintain privacy and access to it *secured*.

Enterprise information governance also includes enforcing data privacy on an enterprise wide basis. This means that data discovery also needs to identify and locate sensitive data within, and across, systems even if that data has been transformed in some way. It should then be the case that data privacy of specific entity data (e.g. customer data, employee data) can be enforced through data masking and run-time screen masking.

INFORMATION GOVERNANCE: COMPONENT TECHNOLOGIES NEEDED

A complete suite of integrated tools is needed to implement information governance

With respect to technology, companies need a standard suite of end-to-end tools in a single integrated information management platform to help them govern core master and transaction data in a standard way. This information management platform is shown below in Figure 4 and includes tools to define and search for data names and definitions of available business data (a business glossary), data modelling, metadata and data discovery, data mapping, data profiling, data cleansing, data integration, data service publishing for provisioning data and data monitoring. This toolset should also allow users to trace where data came from and how it was transformed. This tracing is known as metadata lineage. Sitting on top of the toolset is a data governance management tool that allows you to make use of the underlying tools in the platform to govern a specific type of data e.g. customer master data, product master data or insurance premium transaction data. In that sense

A data management platform is a suite of tools that access shared metadata

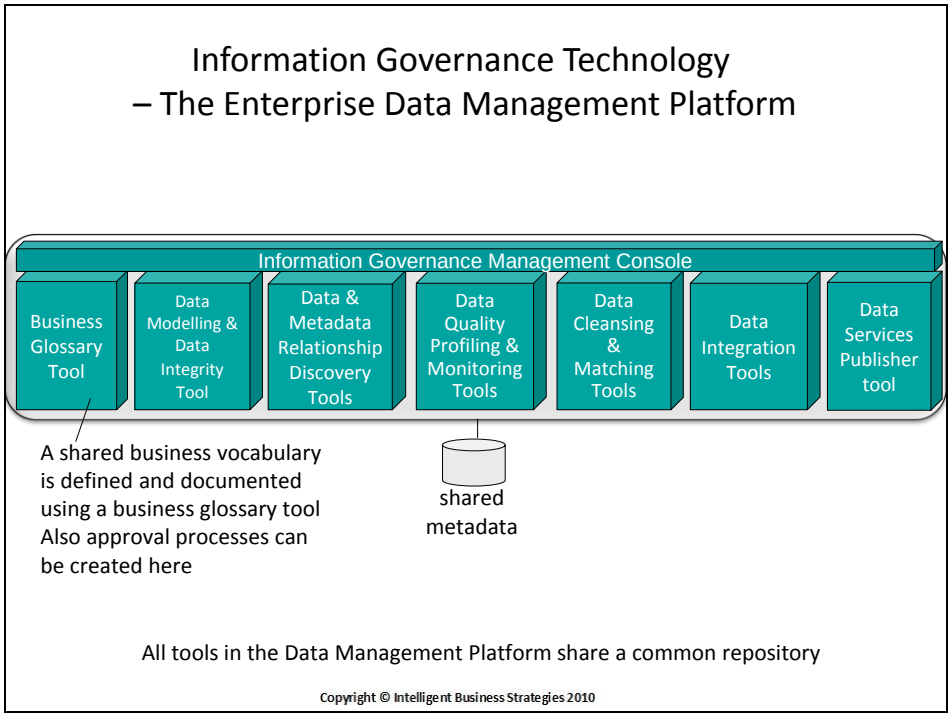

Figure 4

With respect to the steps in the enterprise information governance continuous improvement methodology defined in Figure 3, there should be tools in this integrated enterprise data management platform to support all of the steps in the methodology as indicated in Figure 5.

The business glossary is a focal point for users who need to know what data exists in the enterprise

For example, a business glossary is used to define the data elements associated with master data entities (e.g. asset, product, etc.) and transaction types (e.g. orders) in the shared business vocabulary (SBV) using enterprise wide common data names and definitions. Approval processes should be implemented in this tool to all data item creation/change/de-commission requests to flow to the information governance council. In addition, integration of the business glossary with collaboration and social software would allow enterprise wide participation in defining data items prior to approval. Similarly a data modelling tool is used to design data models using data elements or complete entities taken from the SBV. The data discovery tool can then be used to discover data elements and data relationships in data sources and map these to each other and to the target model before deriving the necessary data transformations needed to integrate and transform the discovered data. Discovered data items and data relationships can then be passed to a data profiling tool to assess quality in all discovered data elements associated with the data entity. Data relationships can also be validated and verified using data profiling, e.g. there are no orders without a customer. Discovered transformation rules and data cleansing rules can then be passed to data integration technology to automatically create data integration jobs that clean and transform the data to get it into a state that is fit for business use.

Data discovery helps to identify where data that needs to be governed is located

Note that this can be done for data warehousing, master data management, data federation and even XML message translation on an enterprise service bus thereby guaranteeing commonly defined high quality data everywhere in the enterprise. Finally data quality and integration web services can be published so that they can easily be integrated with applications and business processes to guarantee that data is consistently validated, integrated and maintained in a common way. It is the sharing of metadata across tools that makes EDG possible.

It should be possible to systematically manage subsets of data

Note that the information governance management console sitting above the other tools in the platform (Figure 3) makes it possible to systematically govern subsets of data e.g. customer master data or orders transaction data. It allows that data to be defined, modelled, discovered, mapped, profiled, cleaned, integrated, provisioned and monitored. It also allows policies to be defined and stewards who govern the data in specific areas of the business to be appointed. These stewards can then enforce data policies, monitor data quality metrics, monitor policy (rule) violations, take actions to remain compliant, track and resolve issues, escalate problems and prioritise the information governance workload.

Each tool in the suite supports a different task in the continuous improvement process

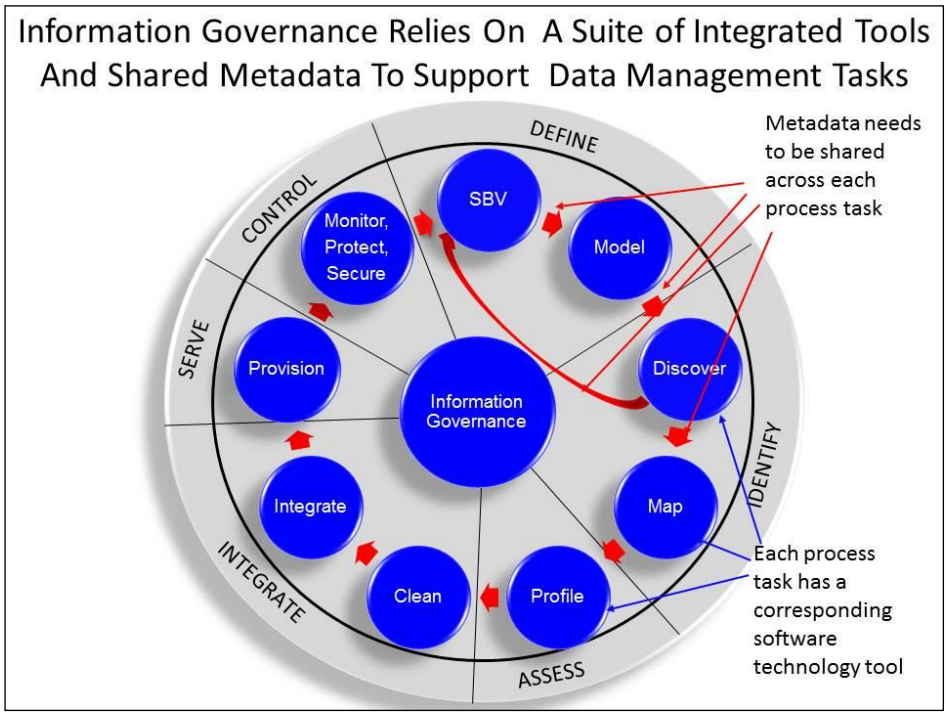

Figure 5

Governing data by master data entity and transaction type makes it easier to understand how information governance is being implemented from a business perspective

The potential here is that by categorising data items to be governed in the SBV by master data entity or transaction-type that the information governance problem can be systematically dismantled and conquered. People, processes and technology can be used to govern and manage each master data entity or transaction-type across the enterprise. Taking the Customer master data entity as an example, Customer information governance would include:

- Customer data definition in the SBV
- Customer data modelling
- Customer data discovery
- Customer data profiling
- Customer data quality
- Customer data integration (federation, consolidation, synchronisation)
- Customer data provisioning
- Customer data monitoring by data stewards

Similarly Customer data management would include

- Customer data backup
- Customer data privacy
- Customer data access security
- Customer data archive
- Customer test data generation

IBM TECHNOLOGIES FOR INFORMATION GOVERNANCE

Having defined what information governance is, why it is needed and what is needed to implement it, this section looks at how one vendor is addressing this problem. That vendor is IBM. IBM provides an integrated suite of tools that together form a platform for end-to-end information management and information governance. These tools included in the platform are as follows:

IBM have a complete suite of integrated tools to implement information governance from end-to-end

IBM Information Management Products	Information Governance Usage
IBM InfoSphere Foundation Tools	Define, model, discover, profile and generate services to establish and govern trusted information and a core component of IBM InfoSphere Information Server
• IBM InfoSphere Business Glossary	Define, manage and control common data names and data definitions for all master data and transaction data that needs to be governed
• IBM InfoSphere Data Architect	Data modelling
• IBM InfoSphere Discovery	Discovery of disparate data within and across source systems that needs to be governed, cleaned and integrated and made fit for business use
• IBM InfoSphere Information Analyzer	Data quality profiling to determine the state of data that needs to be governed and to monitor and make people accountable for data quality to maintain business confidence in it
• IBM InfoSphere FastTrack	Capture design specification mappings and generate data integration services to integrate and clean data
• IBM InfoSphere Metadata Workbench	Monitor data flows - metadata lineage and audit
IBM InfoSphere Information Server	Cleanse, integrate, deliver trusted information from heterogeneous sources
• IBM InfoSphere Blueprint Director	Used to build templates for data warehousing, MDM, data migration, data synchronisation etc. from data quality and data integration services created in underlying IBM InfoSphere Information Server tools as part of an information governance program
• IBM InfoSphere Quality Stage	Data cleansing and matching
• IBM InfoSphere DataStage	Data integration for consolidation
• IBM InfoSphere Federation Server	On-demand data federation to integrate data from multiple underlying data sources
• IBM InfoSphere Services Director	Information service publication for use in managing and governing data
IBM InfoSphere Guardium	Real-time database activity monitoring and database vulnerability assessment
IBM InfoSphere Optim Data Privacy Solution	Data masking for privacy in non-production environments

IBM InfoSphere Optim Test Data Management Solution	Subset data to right-size and speed deployment of testing environments. When combined with masking making them secure
IBM InfoSphere Optim Data Growth Solution	Database archiving
InfoSphere Guardium Data Redaction	Remove sensitive data from unstructured environments (documents, graphics…)
IBM Tivoli Access Manager	Data lifecycle (CRUD) security management

Referring back to the methodology described in Figure 3, these tools can be assigned to each of the steps in the methodology as shown in Figure 6.

Each tool in IBM's InfoSphere information management platform supports a different task in the continuous improvement process

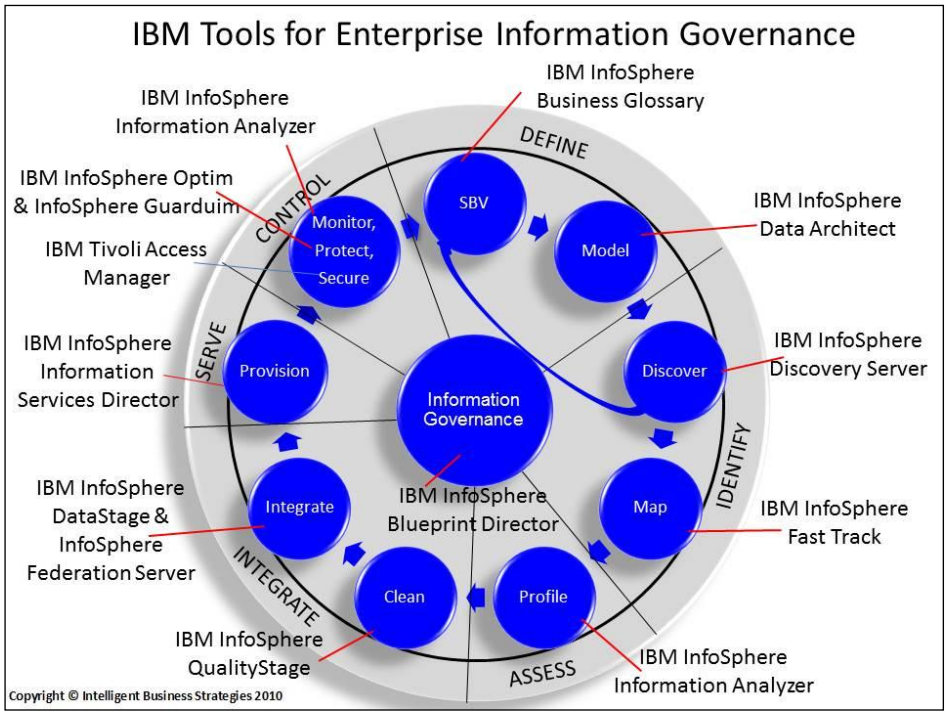

Figure 6

IBM InfoSphere Foundation Tools and IBM'sInfoSphere Information Server products share the same metadata repositiry

InfoSphere Information Server which includes IBM InfoSphere Foundation Tools is at the heart of IBM's solution for structured data governance. Looking at Figure 6, the structured data governance cycle is fully supported by the integrated suite of IBM InfoSphere and InfoSphere Optim solutions.

IBM InfoSphere Business Glossary is a business user tool to define a common set of data names and definitions for data

It can also be used to see what data exists in the enterprise

- IBM InfoSphere Business Glossary is part of the InfoSphere Foundation Tools. This product allows business users to define information to be governed. Data is defined using a common set of data names, definitions and integrity constraints sometimes referred to as a shared business vocabulary (SBV). InfoSphere Business Glossary also supports workflows to control the data naming and definitions process. Business requests for data items to be created, changed or decommissioned can therefore be flowed to an Information Governance Council for approval before being published as data that is used across the enterprise. Users can also search the business glossary to see what data exists in the enterprise and to understand what it means and where it is used. This can be done from a browser or from a mobile device. The latter requires IBM InfoSphere Business Glossary Anywhere mobile software component

- IBM InfoSphere Data Architect is a data modelling tool. Once data has been defined in the business glossary, professional IT data modellers can then make use of the commonly defined data items when building data models by directly accessing the business glossary from within InfoSphere Data Architect. The screenshot in Figure 7 shows IBM InfoSphere Data Architect with the eclipse plug-in of the Business Glossary. Data modellers drag and drop data items in the glossary directly into their data model. This guarantees meaningful business data names in data models and governs the data modelling process. This is an example of how shared metadata improved productivity and governance within the same IBM InfoSphere Information Server platform.

Data modellers can use data items defined in the business glossary when building data models using InfoSphere Data Architect

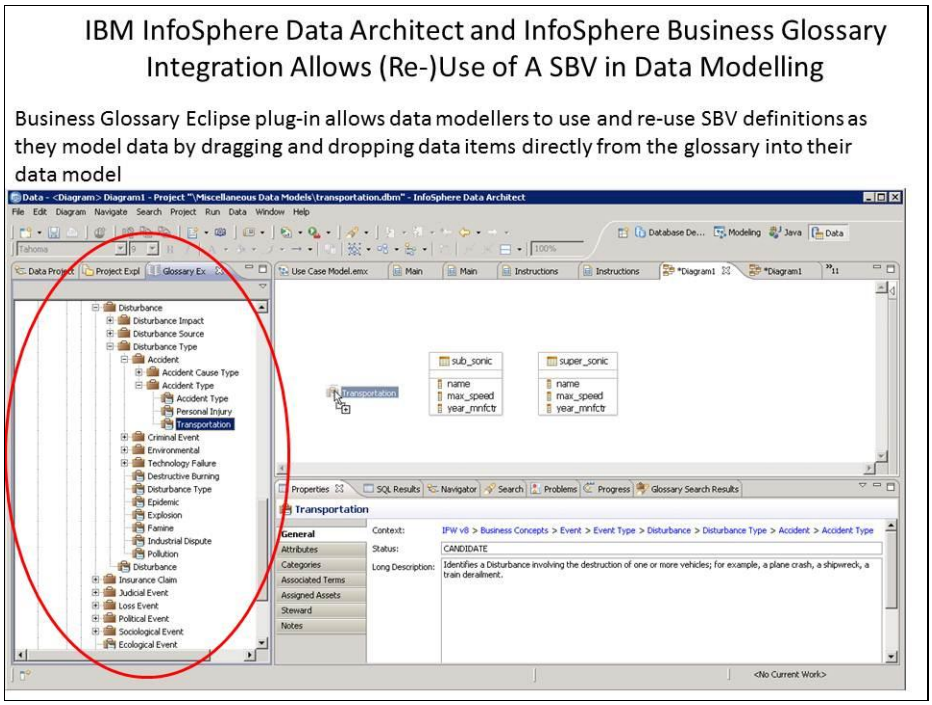

Figure 7

IBM InfoSphere Discovery automatically discovers data and data relationships across multiple source systems to locate the data that needs to be governed

- IBM InfoSphere Discovery is also part of the InfoSphere Foundation Tools suite of products. Using InfoSphere Discovery it becomes possible to *automatically* discover where in the enterprise data is located that needs to be governed. This includes discovery of data relationships within and across systems and also sensitive data that may need to be protected. InfoSphere Discovery uses a cross-profiler, unified schema builder and transformation analyzer to finding complete business data entities (e.g. customer, product, order) as described in Figure 8. Aligning data discovery with these core business entities helps business people understand what data discovery is about and why it is needed.

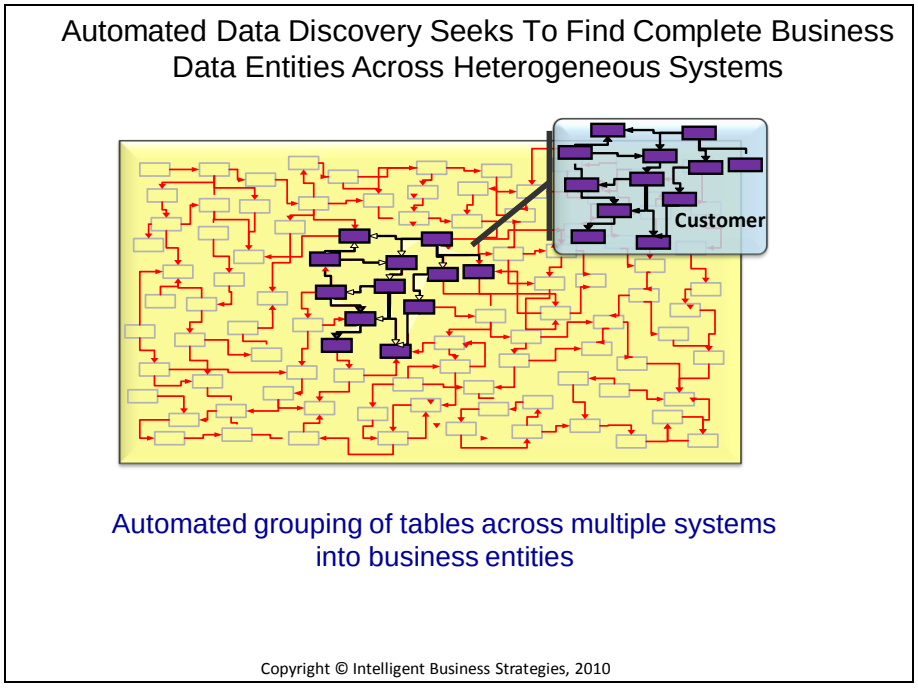

Figure 8

IBM InfoSphere Discovery can automatically determine 'business objects' that can then be governed using other other tools in the platform

InfoSphere Discovery can also pass the metadata it has created on data entities to other tools as shown in Figure 9. This makes it possible to implement information governance at the data entity level irrespective of where the data is located. It is this capability specifically that offers enormous productivity gains in governing information across the enterprise. For example, customer data integration, customer data quality, customer data privacy, customer data retention, customer data access security, and customer data archive are all potentially possible regardless of where the underlying data is stored. IBM InfoSphere Discovery automation can therefore reduce development time by prototyping and testing new transformation rules before data is physically integrated and moved.

IBM InfoSphere Fast Track can automatically generate workflow servces from metadata to integrate and clean data to make it fit for business use

- IBM InfoSphere Fast Track is part of the IBM Information Server platform. This product uses metadata to specify source to target mappings and then *automatically* generates IBM InfoSphere DataStage data integration and IBM InfoSphere QualityStage data cleansing workflow services needed to integrate and clean the discovered data to get it into a form that is 'fit for business use'. Automated generation of data integration and data cleansing workflows significantly reduces the cost and time to develop data integration (also known as Extract, Transform and Load – ETL) and data cleansing services using this agile metadata driven approach. IBM InfoSphere Fast Track can receive mapping metadata from IBM InfoSphere Discovery Server which means that automated discovery can feed metadata to software that automates the generation of data integration services. This capability provides significant development productivity benefits.

IBM InfoSphere Discovery can exchange metadata with IBM InfoSphere Information Server and IBM InfoSphere Optim tools to speed up and automate information governance tasks

Data management and administration tasks can then be performed at the object level regardless of where the underlying data is stored

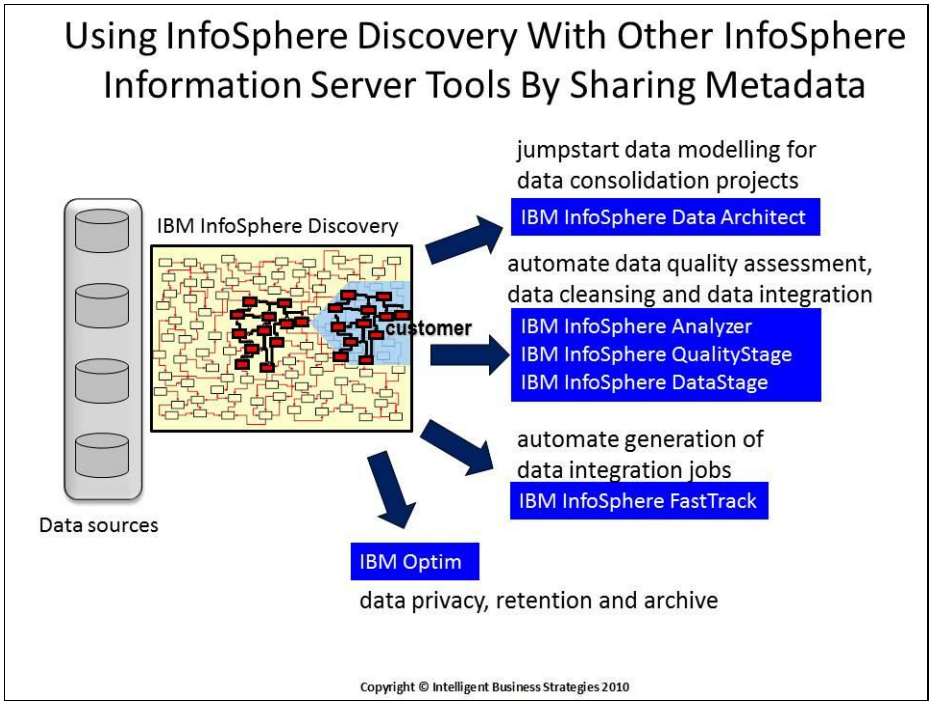

Figure 9

Data can be cleaned and consolidated to build data warehouses, MDM systems and migrate data

- IBM InfoSphere DataStage and QualityStage are used to create data integration and data quality workflows to integrate and cleanse data to get it into a form that is fit for business use. They can be used to build data warehouses and MDM systems as well as to migrate and synchronise data. IBM InfoSphere Fast Track can automatically generate these workflows to facilitate agile and rapid development

Data quality and data integration services can be published so that they can be invoked on-demand to drive consistency everywhere

- IBM InfoSphere Services Director can be used to publish data integration and data quality workflows as web services and make them available on an enterprise service bus (ESB) as part of a service oriented architecture (SOA). In this way data integration and data quality services can be created once and reused to guarantee consistency and quality of data across the enterprise. As published services they can be invoked on-demand from other applications, processes and portals to provision trusted data needed by the business

Data quality can be monitored

- Once data has been provisioned, InfoSphere Information Analyzer can be used by business data stewards to monitor data quality to ensure this does not deteriorate

IBM InfoSphere Optim can mask sensitive data to protect privacy

- IBM InfoSphere Optim can also be used to protect data as part of an information governance initiative. Looking at Figure 9, InfoSphere Optim can receive metadata from InfoSphere Discovery about discovered data and then apply data privacy policies to mask sensitive data that needs to be protected across the enterprise. This includes masking test data. In addition InfoSphere Optim can also manage data retention by archiving data at the data entity level. For example it becomes possible to archive orders data that is over three years old. This capability is particularly useful when data has

to be retained for compliance reasons. InfoSphere Optim allows this type of requirement to be easily managed.

InfoSphere Guardium can monitor and audit database activity across heterogeneous databases in real-time

- Staying with information protection, InfoSphere Guardium provides the capability to proactively identify unauthorised or suspicious activities by continuously tracking all database actions. In addition malicious or unapproved activity by DBAs, developers and outsourced personnel can be detected or blocked without relying on native logs, triggers or other DBMS-resident mechanisms. Pre-configured reports and automated workflows (electronic sign-offs, escalations, etc.) are also available. These make it possible to simplify compliance processes and continuously monitor information protection. InfoSphere Guardium allows organisations to create a continuous, fine-grained audit trail of all database activities, including the "who, what, when, where, and how" of each transaction. It uses agent technology to monitor and audit database activity in real-time across a range of DBMS products that include:
  - Oracle
  - Microsoft SQL Server
  - IBM DB2 (Windows, Unix, z/Linux)
  - IBM DB2 for z/OS
  - IBM DB2 for iSeries (AS/400)
  - IBM Informix
  - MySQL
  - Sybase ASE
  - Sybase IQ
  - Teradata
  - Netezza
  - SharePoint
  - …

Access to information is secured by IBM Tivoli Access Manager

- IBM Tivoli Access Manager is used to administer access to data to help govern the information lifecycle management. This tool can be used with other technologies and applications to manage who is authorised to create, read, update and delete data items

Best pactice blueprints leverage the IBM InfoSphere Information Server platform to govern data

- Last but not least, IBM InfoSphere Blueprint Director sits on top of the InfoSphere Information Server and is used to create best practice blueprints to govern data. So for example, with respect to Customer information governance, several blueprints could be created for:
  - Customer Master Data Management
  - Customer Data Warehousing
  - Customer data migration
  - Customer data quality
  - Customer data replication

  One of the most important things about InfoSphere Blueprint Director can is that it can re-use data integration and data quality services thereby guaranteeing that data is consistent across all information management projects

CONCLUSIONS

Without information governance, process defects and poor decision making will occur which can damage overall business performance, increase risk and cause compliance violations

We have seen in this paper that business intelligence, risk management and compliance are now heavily dependent on trusted, protected and secure data and as such are key business drivers for embarking on an information governance project. In addition, the eradication of data defects also helps improve efficiency of operational business processes.

However, as companies create and consume increasing amounts of information from more and more internal and external data sources the challenge of governing structured and unstructured information is becoming increasingly difficult. In order to get information under control, companies need to organise themselves and define policies and processes needed to govern information across the enterprise. A continuous improvement methodology is also needed.

Companies need an integrated suite of tools with shared metadata to get their information under control

It is already the case that stand-alone tools supporting specific data management disciplines are no longer good enough. What is needed now is a complete integrated suite of tools for end-to-end information governance that share metadata and that can be used together to systematically get information under control.

There is no question that IBM has understood the demand for this in the marketplace. They have invested heavily in information management over the last several years and have produced an integrated suite of tools in IBM InfoSphere Information Server and IBM InfoSphere Foundation Tools that lead the market in end-to-end structured information governance. In addition they have also integrated this platform with IBM MDM Server and IBM Cognos 10. This, coupled with shared metadata across the entire platform and full metadata lineage, makes IBM a certain shortlist contender for those companies trying to get their information under control.

ABOUT INTELLIGENT BUSINESS STRATEGIES

Today, successful companies are those that can absorb new information technologies and use them effectively in their businesses. But faced with so many new technology developments how can IT and business users possibly keep up? Intelligent Business Strategies is an IT research and consulting company whose goal is to help companies understand and exploit new developments in business intelligence, analytical processing, data management and enterprise business integration. Together, these technologies help an organisation become an *intelligent business*.

Mike Ferguson is Managing Director of Intelligent Business Strategies Limited. As an analyst and consultant he specialises in business intelligence and enterprise business integration. With over 29 years of IT experience, Mike has consulted for dozens of companies on business intelligence, data management and enterprise business integration. He has spoken at events all over the world and written numerous articles. Mike is a resident expert on the B-EYE-Network, providing articles, blogs and his insights on the industry. Formerly he was a principal and co-founder of Codd and Date Europe Limited – the inventors of the Relational Model, a Chief Architect at Teradata and European Managing Director of Database Associates. He teaches popular master classes in Operational Business Intelligence and Performance Management, Enterprise Data Governance, Master Data Management and Enterprise Business Integration

INTELLIGENT BUSINESS STRATEGIES

Intelligent Business Strategies
Springfield House
Water Lane, Wilmslow
Cheshire SK9 5BG
England
Telephone: (+44)-1625-520700

Internet URL: www.intelligentbusiness.biz
E-mail: mferguson@intelligentbusiness.biz

Succeeding With Information Governance Using IBM Technologies