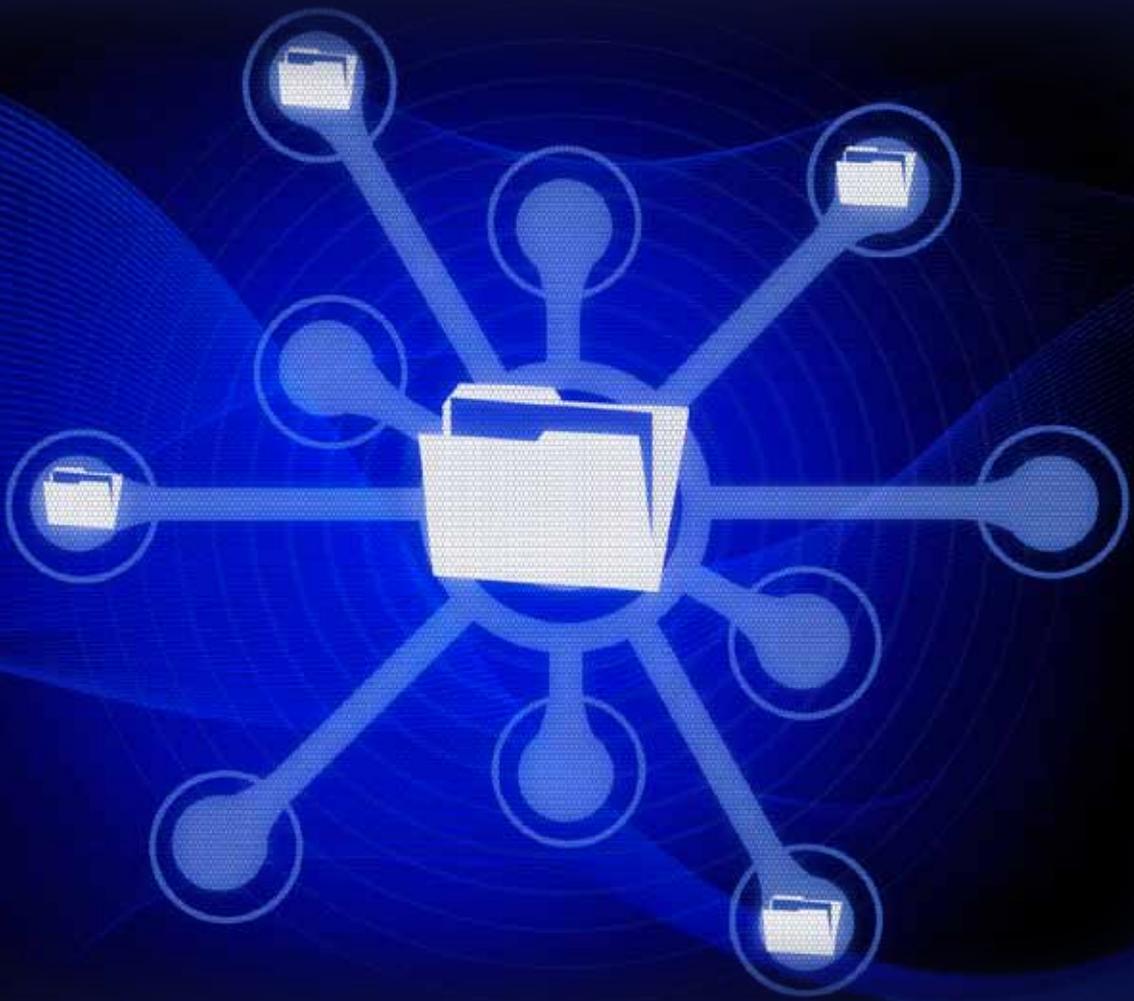# Data
Management

Arkansas Approach

Executive Summary

"Data is the oil of the 21st century". These succinct words by Peter Sondergaard, head of Gartner Research, reflects the growing importance accorded to data[1]. Industry has come to understand that data will be the resource running the economy in years to come.

Arkansas embraced information technology in the early years of electronic computing and, along the way, earned accolades with examples of excellent executions such as e-governance through Arkansas.gov, innovative child support systems etc. Over the years, the state has developed a vast network of computing resources. Data, generated within the state's computing resources and networks, is an asset of the state. As an asset it should be managed efficiently, administered and shared consistently with the requirements of security, privacy, and confidentiality, with timely access to state officials in useful forms [ACA § 25-4-102 (a)(1)(3)(5)]. Like any other asset, data needs care, feeding, and careful management. This Data Management Strategy document is aligned with Arkansas legislation regarding data.

Introduction

Data management is a relatively new discipline when compared to traditional asset management disciplines such as financial management and capital management. The state should lead in this rapidly developing critical area. The state of Arkansas has chosen the Data Management Association Data Management Body of Knowledge (DAMA DMBOK) framework for the management of this new asset type. The DAMA framework was chosen because it is consensus driven by the largest body of data professionals worldwide not affiliated with any specific vendor or technology. The data management terms and definitions in this document align with DAMA DMBOK. A common vocabulary in the data management domain is important for this new discipline, and DAMA, through its non-profit, consensus driven approach has created the most acceptable version of the glossary in the industry. Taking advantage of existing DAMA definitions, a glossary has not been appended to this document.

Besides establishing a common language for data management, the DAMA DMBOK provides a data management framework that is holistic and covers all sectors within data management (shown in Figure 1). Every organization is unique and all sectors may not have equal relevance within organizations if at all. The prioritization and depth of focus on sectors is for organizations to decide.

---

1. Data and information are synonymous and have been used interchangeably within this document
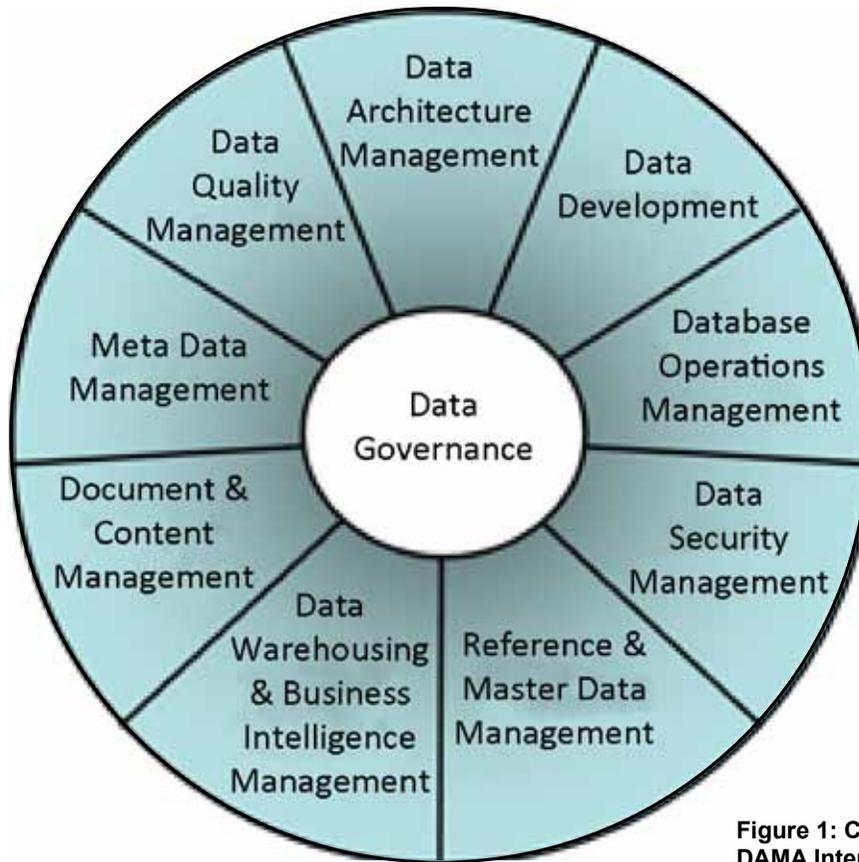
**Figure 1: Copyright
DAMA International 2010**

Business Case: The value proposition for data management strategies has clear benefits. Applying sound data management practices and standards leads to clear, documented, effective data, and reduces attendant liabilities with sensitive data. Better quality data results in cost savings in state government operations.

In addition to cost savings and reducing risk, data management can improve the delivery of services to Arkansas citizens by:

Ensuring the linkage of data resources to legislated mandates and state goals
- Improving interoperability and integration of systems
- Increasing organizational flexibility and agility to meet changing requirements
- Identifying innovation opportunities

Without data management strategies defined at the enterprise level,
- Common issues are addressed differently, if at all
- Lack of one common strategy results in costly reinvention of the best practices, policies, and solutions
- Individual and unplanned approaches can lead to less desirable results

Vision Statement

Excelling in data management for the good of the citizens of Arkansas, guided by carefully crafted data management strategy to maximize data value and minimize data risks.

Implementation:

The data management strategies in this document have to be ratified by a data governance body before publication as best practices for the state of Arkansas. The data governance body may choose to modify this document in any manner it deems appropriate at different review periods. It is proposed that this document be reviewed biennially by the data governance body. The data governance body may choose to audit some of the mandates within this document and act upon observed discrepancies.

Methodology of Strategy Development:

The data management strategies were created based on Arkansas data principles and guided by legislation (ACA 25-4-102 (a)) and Arkansas goals. The strategies have been classified and enriched using the DAMA DMBOK framework.

State of Arkansas Legislation [ACA § 25-4-102 (a)(1),(3),(5)]

(a)   The General Assembly finds and declares information and information resources to be strategic assets of the state of Arkansas and that procedures must be established to ensure that:
(1) Information resources are used in an efficient manner;
(3) Information is administered and shared consistently with requirements for security, privacy, and confidentiality;
(5) State officials have timely access to information in useful forms;

State of Arkansas Goals:

1. Improve education for Arkansans
2. Improve economic development
3. Increase efficiency in state government
4. Improve citizen access to state services
5. Protect the environment

Arkansas Data Principles:

1. Manage enterprise data as a state asset
2. Enable openness and transparency
3. Share data to enhance its value
4. Enforce privacy and security
5. Integrate common data definitions and standards
6. Collaborate to eliminate duplicates
7. Improve state government through data quality

1. Data Governance:

Definition: *Data governance is the execution of authority and control (planning, monitoring and enforcement) over the management of data assets.*

Data governance impacts all areas of data management and directly influences and prioritizes the data management strategies within this document. It is important to distinguish data governance from IT governance; it is different in that it is somewhere between business and IT governance. For instance, Health Insurance Portability and Accountability Act (HIPAA) compliance involves both business and IT participation. In organizations, data needs are framed by the business and should be audited by the business for compliance and quality while IT implements and operates the infrastructure for the data. Data governance needs to be a partnership that includes business stewards who decide on the use and control of the data and technology stewards who enable and administer the flow and storage of the data. The business stewards are trustees of the data while the technology stewards are custodians of the data. The business and technology stewards are not new jobs but a formalization of existing roles within different agencies where data governance would enable shared decision making about data assets. Data issue management, where difficult decisions need to be made, is a key activity of data governance.

Data governance often deals with data usage and its legal implications. Legal counsel is advisable within the highest data governance body to interpret laws and attempt changes, if necessary, for the greater good of the citizens of Arkansas. Ensuring compliance with the laws involving data is an essential part of data governance, making auditing an essential component of the Data Governance Council.

Strategy: *Obtain highest possible executive level support at the state level for data governance.*

Strategy: *Educate about the need for data governance.*

Strategy: *Form working group to create the decision rights and accountability structures for a Data Governance Council.*

Strategy: *Have a top level State Governance Council primarily representing non IT leadership with very few meetings per year but provide directions to data governance steering committees to work on the details.*
Strategy: *Develop a state data governance charter based on collaboration, mutual support, and transparency.*

Strategy: *Include representation from Arkansas Legislative Audit and Arkansas Attorney General's Office.*

Strategy: *Form a data governance body to provide staff support to data governance function, facilitate meetings, prepare meeting agendas, and publish minutes.*

2. Data Architecture Management:

Definition: *Defining the data needs of an organization and designing the master blue prints to meet those needs.*

Data architecture is an integrated set of specification documents that describe the information needs of an organization, guides integration and control of data assets, and aligns data investments with data strategy. As the data needs of an organization becomes more complex, the value of data architecture increases by helping to understand how different pieces work together at various levels of abstraction. The artifacts could be tables, charts, diagrams, documents, or data models. Depending upon business need, artifacts may be either very brief or extremely detailed. Specific outputs of data architecture include but are not limited to the enterprise data model, information value chain analysis (relationship between data and business processes, roles and organizations etc.), data integration (aka data flow) diagrams, data technology architectures.

Strategy: *Identify communities of interest (COI) within state government by subject or purpose (example health care or fraud prevention) that would benefit most from enterprise data models, standardized data exchange schemas and namespaces.*

Strategy: *When contemplating data exchange models (such as in SOA XML schemas), take advantage of established industry standards if any (e.g. HL7 in health care) or federal data standards (e.g. NIEM in law enforcement and justice) as most appropriate for the business area.*

Strategy: *When designing data exchange models apply the principle of "open to extension and closed to change."*

Strategy: *Define and maintain data integration relevant to identified COIs with a series of matrices mapping the relationship of data with business roles (CRUD), organizations, business function, locations etc.*

Strategy: *DIS enterprise architecture team to inventory and maintain data technology architecture interviewing cross-section of state organizations, large and small.*

Strategy: *DIS data architecture team should be familiar with architectural frameworks such as The Open Group Architecture Framework (TOGAF), Zachman, and Federal Enterprise Architecture (FEA) Data Reference Model etc. to understand and apply architectural best practices in alignment with enterprise architecture.*

3. Data Development:

Definition: *Designing, implementing and maintaining the solutions to meet the data needs of an organization.*

The practice of analyzing, designing, implementing and maintaining data products for an organization is data development. The end data products are data models, physical data structures, and information end products such as screens and reports, all with the aim to support a range of business activities from strategy development to operations. Data development activities may include data architects, solution architects, business analysts, data analysts, software developers, business stewards

and business subject matter experts (SMEs), all working together to produce the data products. Depending upon the project and organization size, one or more of these roles may be the function of one individual. Data development touches various phases of the system development lifecycle (SDLC) where data is defined, designed, and implemented, whether in the traditional waterfall method or the shortened phases of agile methodologies.

Strategy: Invest in enterprise agreement on the business definitions for critical data elements in the early phases of requirements gathering, towards a business glossary.

Strategy: Mandate data models for any IT solutions development.

Strategy: Mandate model change policy and incorporate model change processes under change management processes. Mandate synchronization between models and databases.

Strategy: Develop entity, attribute and table, and column naming standards and conventions.

Strategy: Data should reflect actual entities and attributes of the business and not be tied to a specific application. Implement application specific requirements through data virtualization using views, stored procedures and functions.

Strategy: Database processing should be pushed towards the database server by design rather than the application server.

Strategy: Enforce data rules closer to the database when possible, rather than in application code.

Strategy: Develop test data meeting privacy and confidentiality requirements.

Strategy: Consider implementing SQL review practices between developer and DBA functions to prevent production failures, enhance performance and enhance maintainability.

Strategy: Automate, to the maximum extent, the data migration from source to target platforms early in the development phases.

Strategy: Have policy not to update production data directly though ad-hoc updates.

Strategy: Cross train, as appropriate, application DBAs in non-relational technologies such as XML, XML Schema, Namespaces and OO developers in SQL best practices using relational databases.

4. Database Operations Management:

Definition: Planning, monitoring, control and support of structured data assets across the data assets lifecycle.

Database operations management is among the most mature of data management areas with the best practices tested over decades and refined by large networks of professionals, primarily database administrators. Database operations management

Arkansas Data Management

covers two main areas a) database support and b) data technology management. Database administrators in coordination with other IT functions attempt to maximize the value of structured data assets in the organization by a) protecting and ensuring the integrity of the data, b) maximizing availability of the data and c) optimizing database performance. These goals are supported through many activities such as:

1. Backup and recovery planning and management
2. Database monitoring and tuning
3. Ensuring appropriate versions of database technologies are being used
4. Running various data operations such as loading, reorganizing databases, data statistics refresh, archival and purging
5. Evaluating new data technologies appropriate to the organization

Database administration, besides being central to the database operations function, plays important roles in other data management areas such as data development and data security management.

> **Strategy:** Data archival policy and standards should be developed and followed to avoid overloading of production databases leading to performance degradation over time.

> **Strategy:** The data purge policy in alignment with the Arkansas Records Retention Schedule and the needs of the business should be developed and followed. This is not to be confused with the archival policy as being the same since archival and purging are two separate activities.

> **Strategy:** Organizations should verify the validity of its backups through recovery exercises at least once a year.

> **Strategy:** While developing applications for problem monitoring, invest in management by exception, rather than reporting messages of success, to prevent eventual information overload with inefficiencies and/or inefficacies.

> **Strategy:** Production database change policy should always mandate a documented back-out plan for every change.

> **Strategy:** Have policy to always test changes in test environments with the exception of emergencies

> **Strategy:** Have policy to develop automation skills within DBA community.

> **Strategy:** Database de-normalization should be among the least preferred performance strategy within online transaction processing (OLTP) databases.

> **Strategy:** Invest in the practice of proof-of-concept activities for new and promising technologies to build roster of suitable technologies in advance. This would help to avoid overestimation of benefits & underestimation of costs when implementation opportunities surface.

> **Strategy:** Decide and document database management software upgrade policy even if the policy is limited to reacting to vendor end-of-support ultimatums. A documented upgrade policy is in the interest of better infrastructure resource planning.

5. Data Security Management:

Definition: *Planning, development, and execution of data security policies and procedures to provide proper authentication, authorization, access and auditing of data and information.*

It is not the organization size, but the business nature that dictates the effort needed for data security management. Organizations dealing with sensitive personal information would need to invest more than others for date security management. A proper balance between data access and data security should be maintained. Sweeping iron clad security policies may stifle beneficial uses of data and generate resentment within an organization. Data security, carefully managed with monitoring, auditing, and enforcement promotes trust amongst stakeholders. This trust encourages data sharing and thereby increases data value. Organizations will be reluctant to share information unless appropriate security stewardship of data can be assured. Data security should have judicious governance with stakeholders so that it is practical to be followed daily on an operational level.

Trends toward cloud computing bring special data security concerns. Organizations can move data and associated security controls, but not liability, to the cloud. Special attention should be paid to data moving to the cloud and the contractual content with the cloud vendor.

Strategy: *Organizations should have the Arkansas mandated SS-70-001 "Data and System Security Classification" completed and reviewed biannually for needed changes (new solutions, legislations, rules etc.).*

Strategy: *Based on data security classifications, organizations need to address sensitive data exposed in test databases though data masking or de-identification.*

Strategy: *Develop data sharing agreement templates that organizations could leverage when crafting interagency data sharing.*
Strategy: *Organizations should follow password management requirements as mandated in SS-70-002 "User Logon Passwords".*

Strategy: *Organizations with sensitive data should manage, at some level, a log of access granted to roles and individuals.*

Strategy: *Access to sensitive data should be avoided through shared accounts.*
Strategy: *Manage access through role based security at the group level rather than individual based accounts. Assign individuals to roles.*

Strategy: *Grant access to sensitive data through approved and not through default opt-in.*

Strategy: *For very sensitive information, provide for authentication and access monitoring of unusual patterns with a judicial balance of automation and human checks.*

Strategy: *Incorporate annual auditing processes not framed with a fault finding mindset but with an objective to monitor for continuous improvement.*

Strategy: *Develop a cloud based data security strategy.*

6. Reference & Master Data Management:

Definition: *Planning, implementation, and control activities to ensure consistency with a "golden version" of contextual data values.*

Every business transaction record needs context. For example, when a customer places an order for a certain quantity of products, at a certain price, the customer, product, and order status are contextual data while the order quantity, discount, and price are transaction data. Organizations are facing challenges in keeping contextual data consistent across lines of businesses and systems. Contextual data maintained in silos make organization integration difficult with the inevitable inconsistencies. The root cause analysis of many data quality issues within organizations points to the need for master and reference data integration. The overall data quality in many organizations is directly correlated with the quality of the contextual data. Reference and master data management are essentially data quality programs at higher levels of the organization.

There are two kinds of contextual data, reference data and master data. In the example above, the customer and product information are master data while order status is the reference data. Reference data commonly appears as a pick list within applications. Reference data categorizes data for business purposes and, therefore, the domain values have to be controlled with definitions for each value and with its relationship with other values with the domain. Master data, once defined at the entity level, does not require every element defined. The challenge with master data, however, is prevention of duplicates and the creation of a "golden" record with the merger or most accurate elements from disparate sources, and the subsequent dissemination of master data. Governance structures are essential for reference and master data management projects because data conflicts cannot always be resolved through automation and established procedures.

Strategy: *Identify possible COIs within state government that may benefit from master data management (MDM) efforts.*

Strategy: *Develop robust metadata, including business glossary, at the beginning of an MDM effort versus documentation at the end.*
Strategy: *Plan for data governance as a must - not optional, when approaching a reference or master data management project.*

Strategy: *Invest in master and reference data management efforts as a continual program and not as a project with an end date.*

Strategy: *Invest in master and reference data management efforts in smaller iterations to deliver and demonstrate value and continued support from stakeholders.*

7. Data Warehousing & Business Intelligence Management:

Definition: *Data warehousing and business intelligence management (DW-BIM) covers the planning, implementation and control activities in the gathering, cleansing, integration and presentation of data to knowledge workers for business analysis, thereby enabling informed decision making by organizations.*

Data warehousing is the activity that is concerned with the collection of data from various data sources within the organization, integrating it and storing it as a snapshot of organizational operations at different points in time. In other words, the concern is about integrated enterprise data content with a historical perspective. BIM is the complementary part of using this data content using various tools. These two activities are intertwined in that one is ineffective without quality management in the other.

Within state government, the communities sharing common data interests (interagency) or even within an agency need to identify information they would benefit from measuring and tracking over time. Pooling knowledge analysis skills, technical skills and software in the DW-BIM domain would help lower costs and increase quality to move toward a DW-BIM Center of Excellence within state government.

The term "big data" is getting a lot of press during the time this strategy document was being drafted in 2013. Big Data refers to data whose variety, volume and velocity exceeds the capabilities of conventional database technologies. The literature of the times pointed toward the need to invest early in Big Data for competitive advantage. However, the DIS Data Architecture team considers "Big Data" a new name for specialized data mining research and operations of the past dealing with streaming data or other data too large to be captured and then analyzed. The size of Arkansas state government and budget may not warrant the state being an early adopter of these new technologies.

Strategy: *Leverage and support data management component functions such as reference and master data management, data governance, data quality and metadata management.*

Strategy: *Develop state government data warehousing and business intelligence technical architecture (inventory).*

Strategy: *At the minimum, for any incremental addition to the data warehouse, develop executive summary with business review and buy-in. Business sponsors, business case, scope and resources are to be identified at the start of any DW-BIM initiatives and never at later stages.*

Strategy: *In the interest of enterprise integration, allocate substantial time to explore re-use, sharing and extension of existing dimensions when adding new business fact tables to cover additional business processes.*

Strategy: *Summarize and optimize last, not first. Start building with the detailed data.*

Strategy: *For "big data" situations, seek and collaborate with the data mining faculty and researchers of the state's leading universities.*

Strategy: *At the state government level, avoid investing in "big data" as early adopters and instead annually review use cases for appropriateness.*

Strategy: *Actively support and invest in metadata policy and processes within the organization with the business glossary process among the initial steps.*

Strategy: *Actively support and invest in enterprise data governance efforts.*

8. Document & Content Management:

Definition: *Document and content management are the planning, implementation and control activities to store, protect, and access unstructured data within electronic files and physical records that include text, graphics, images, audio and video.*

This area refers to unstructured data that is not in the structured format of traditional data management systems (relational, hierarchical, object, networked etc.). Though "store, protect and access" activities within document and content management may seem to imply an operational focus, it is very important to consider strategic aspects of data governance, architecture, security, privacy and confidentiality, metadata and classification, and data quality. Document management is more to do with storage, inventory and control of paper or electronic documents using processes and technologies whereas content management refers to processes and technologies that are concerned with the organization, categorization and access to the content within those documents and records. Content management today is particularly important in managing content within web sites and portals. Document and content management, though distinct, are, in practice, sometimes blurred with business process and roles intertwining and vendors providing products that cover both areas. This is reflected in the Department of Information Systems' enterprise content management strategy and video strategy documents which provide further details on the strategies mentioned within this document (available from the DIS Enterprise Architecture group upon request).

Strategy: *Refer to "enterprise content management" strategy and "video" strategy documents developed by the Department of Information Systems (available upon request from the department's enterprise architecture group).*

9. Metadata Management:

Definition: *Metadata management is the set of processes that ensure proper creation, storage, integration, and control to support the associated usage of metadata.*

Lack of metadata is the bane of organizations large and small. The lack of meaningful and maintained metadata leads to inefficiencies such as 1) higher retraining costs with labor/vendor turnover 2) higher time-to-market for solutions and system changes 3) more time spent in research by data analysts validating or reporting data 4) incorrect business decisions based on lack of understanding of data, 5) lack of understanding between business and IT. For instance, metadata often belongs to the deferred wish-list of application managers maintaining solutions but becomes a must-have during major changes.

Metadata is more than the data dictionary extracted from physical databases or models in a data modeling tool. It is an amalgamation of technical and business understanding of what data is required for the organization to function. There are no boundaries dictating the "right amount" of metadata and it all depends case by case. The amount of technical and business information about data elements should be proportional to its importance within an organization. Metadata may be comprised of business, operational, technical, process or stewardship metadata.

Independent organizations may not have the resources to invest in researching and implementing best practices, policies and procedures in this area. This is one area where collaborative work may be of most help, statewide. It may be noted, however, that though organizations have recognized the importance of maintaining metadata, the success rate, historically, is low indicating that it may be a difficult program to implement.

Strategy: *A metadata group to develop the metadata strategy should be among the first areas addressed though data governance.*

Strategy: *Focus on the governance of metadata toward high quality metadata, the most important aspect for the success of any metadata program.*

Strategy: *Start small (but scalable) at the local level with the most critical business elements.*

Strategy: *For every effort, articulate the problem and/or risk driving the metadata management effort.*

Strategy: *Explore suitable tools for metadata management whether in-house or commercial, for ease of integration, accessibility, and maintainability of metadata.*

10. Data Quality Management:

Definition: *Data quality management is the planning, implementation, and control activities that apply data quality management techniques to measure, assess, improve, and ensure the fitness of data for use.*

Central to the concept of data quality management is the specification of the data needs, determination of the optimal methods to measure and monitor it, agreement of acceptable levels and root cause corrections when there is a deviation from the acceptable levels. The threshold of acceptable quality for the business is to be carefully determined and not be pegged at a level so stringent that is too costly and hence not viable for the organization. Data quality management is not a one-time effort but a continuous program of monitoring and corrections. With a goal of continuous improvement, the acceptable threshold of data quality should always be a moving target. The surfacing of master and reference data management initiatives within organizations has furthered the need for data quality management and the usage of COTS data quality tools. The state should promote the awareness of data quality management and the tools that help in the process.

Strategy: *Develop and maintain inventory of data quality tools in the state with usage licensing and cost.*

Strategy: *Promote awareness by educating on the data quality tool functions, state government success stories, the need for address standardization and quality.*

Strategy: *Seek to arrest the proliferation of multiple vendor offerings in the interest of reducing overall costs.*

Strategy: *Whenever possible, follow industry and federal data standards.*