Metadata Management as a Key Component to Data Governance, Data Stewardship, and Data Quality Management

Wednesday, July 20th 2016
Today’s Presenter: Dalton Cervo

Dalton Cervo has over 26 years of experience in data management, project management, and software development, including architecture design and implementation of multiple MDM solutions; and management of data quality, data integration, metadata, data governance, and data stewardship programs.

Dalton is a member of the Data Quality PRO expert panel, has served on customer advisory boards, and is an active contributor to the data management community through conferences and social media vehicles (blog, twitter).

dcervo@datasourceconsulting.com
/dcervo
@dcervo
Multi-Domain Master Data Management delivers practical guidance and specific instructions to help guide planners and practitioners through the challenges of a multi-domain master data management (MDM) implementation. Authors Mark Allen and Dalton Cervo bring their expertise to you in the only reference you need to help your organization take master data management to the next level by incorporating it across multiple domains.

Morgan Kaufmann, Elsevier, April 2015

Authors Dalton Cervo and Mark Allen show you how to implement Master Data Management (MDM) within your business model to create a more quality controlled approach. Focusing on techniques that can improve data quality management, lower data maintenance costs, reduce corporate and compliance risks, and drive increased efficiency in customer data management practices, the book will guide you in successfully managing and maintaining your customer master data.

John Wiley & Sons, May 2011
Agenda

• A Customer’s Story
• Metadata and Metadata Management
• Metadata Management / Governance
• Metadata Management / Stewardship
• Metadata Management / Data Quality Management
• Conclusion
A CUSTOMER’S STORY
A Customer’s Story

• Problem: what’s the enterprise definition of Net Charge Off (NCO), how is it calculated, and who uses it?
• Generic definition of Net Charge Off (from Investopedia):
  – A net charge off (NCO) is the dollar amount representing the difference between gross charge-offs and any subsequent recoveries of delinquent debt. Net charge offs refer to debt owed to a company that is unlikely to be recovered by that company. This "bad debt" often written off and classified as gross charge-offs. If, at a later date, some money is recovered on the debt, the amount is subtracted from the gross charge-offs to compute the net charge-off value.
• Challenges:
  – The many interpretations by: accounting, collections, risk, corporate planning, and remarketing
  – Identify sources of relevant data
  – Determine calculations utilized
  – Estimate actual value of the asset
  – How to compute expenses and recoveries after charge off
  – Impact of accounting time period
METADATA AND METADATA MANAGEMENT
What’s Metadata?
It’s more than just “data about data”

“Metadata is structured information that describes, explains, locates, or otherwise makes it easier to retrieve, use, or manage an information resource.”

NISO – National Information Standards Organization
Metadata Categories

**Business Metadata**
- Business definitions
- Business rules, regulations, and data quality expectations

**Technical Metadata**
- Physical data structures and interfaces
- Documentation for auditing derivations, dependencies, and data flow

**Operational Metadata**
- Statistics about data movement: frequency, record counts, component by component analysis and other statistics
Islands of Metadata

- Correspondence & Legal Documents
- Business Definitions and Processes
- Business and Data Quality Rules
- Structured Data Sources
- Semi-Structured Data Sources
- Unstructured Data Sources
- Interfaces and Transformations
- Business Applications
- Data Models
- Analytics & Reporting
Metadata Across the Organization

Metadata Repository

- Enterprise Business Glossary
- Enterprise Process Glossary
- LOB’s Business Glossary
- LOB’s Business Process Glossary
- Conceptual Models
- Logical Models
- Physical Models
- Physical Structures
- Data Dictionary
- Data Lineage
- Interface Information
- Data Transformations
- Batch Job Descriptions
- Data Movement Statistics
- Data Security Rules
- Report and Correspondence Mapping
Metadata Management

COLLECT

Metadata Repository

MANAGE

DISTRIBUTE
Collecting Metadata

- Business Track
- Technical Track

- Business Metadata
- Technical Metadata
- Operational Metadata
Distributing Metadata

- Technical Metadata
- Business Metadata
- Operational Metadata

Business Track

Technical Track
Managing Metadata

Organizing, Categorizing, Approving, Maintaining, & Facilitating

- Data Lifecycle Management
- Data Ownership Management
- Data Quality Rule Management
- Data Security Management
- Data Transformation
- Impact Analysis
- Audit Trail
- Risk Mitigation
- Business Rule Management
- Enterprise Business Glossary
Technology in Metadata Management

- Why is technology important?
- Considerations when selecting a vendor
- Business/IT considerations
- Adoption, implementation, and maintenance challenges

- Some of the players:

  - collibra
  - Data Advantage Group
  - SAS
  - IBM
  - Informatica
  - Information Builders
Metadata Management &

DATA GOVERNANCE
There are varying perspectives on Data Governance…

The exercise of **authority**, **control** and shared decision-making (planning, monitoring and enforcement) over the **management** of data assets. Data Governance is high-level planning and **control** over data **management**. **DAMA**

...unites people, **process**, and technology to change the way data assets are acquired, **managed**, maintained, transformed into information, shared across the company as common knowledge, and consistently leveraged by the business to improve profitability. **Chris Deger**

...is a **control** that ensures that the data entry by an operations team member or by an automated process meets precise **standards**, such as a business rule, a **data definition** and data integrity constraints in the data model. **Wikipedia**

...is a system of decision rights and **accountabilities** for information-related processes, executed according to agreed-upon models which describe who can take what actions with what information, and when, under what circumstances, using what methods. **DG Institute**

...the specification of decision rights and an **accountability** framework to ensure appropriate behavior in the valuation, creation, storage, use, archiving and deletion of information. It includes the **processes**, roles and policies, **standards** and metrics that ensure the effective and efficient use of information in enabling an organization to achieve its goals. **Gartner**

...refers to the overall **management** of the availability, usability, integrity, and security of the data employed in an enterprise. **Tech Target**

The execution and enforcement of **authority** over the **management** of data assets and the performance of data functions. **TDAN**
The Datasource Definition of Data Governance (DG)

- Datasource uses Data Governance as an umbrella concept to cover the disciplines often referred to as Data Governance (DG) and Data Management (DM).

- From a DG perspective, it defines who in the organization gets to make what decisions about what data and establishes process and structure to support that governance.

- From a DM perspective, it facilitates and coordinates the myriad of enterprise functions and organizations, processes and technologies to bring about data value optimization.

- Where there are gaps in realizing data value optimization, Data Governance works with the organization to fill them.
Ineffective Interdepartmental Communication
- Conceals reliance on common data
- Ramps up redundant work
- Increases shadow IT data costs
- Magnifies management confusion

Data Quality Issues & Perceptions
- Breaks systems & processes
- Impacts analytics & reporting
- Tramples data trust
- Atrophies operational agility
- Breeds bad business decisions

Deficient Data Change Control
- Breaks systems & processes
- Impacts analytics & reporting
- Tramples data trust

Insufficient Integration / Desktop Integration
- Increases data system costs
- Increases shadow IT data costs
- Atrophies operational agility
- Hampers enterprise perspective

Information Security Confusion
- Compounds compliance risk
- Raises data security risk
- Hampers data access
- Limits performance management prospects

Battling Business Rules
- Increases compliance risk
- Tramples data trust
- Creates reports / analysis conflicts
- Ramps up redundant work
- Magnifies management confusion
Diminishing Value of Your Data, Today

7 Analytic Group Fragmentation & Regulatory Disconnects
   • Creates reports / analysis conflicts
   • Magnifies management confusion
   • Compounds compliance risk
   • Tramples data trust
   • Ramps up redundant work
   • Limits performance management prospects

8 Inadequate Data Accountability
   • Complicates communication
   • Compounds compliance risk
   • Raises data security risk
   • Magnifies management confusion
   • Promotes non-productive work
   • Atrophies operational agility
   • Tramples data trust
   • Conceals reliance on common data

9 Lack of Common Names, Definitions, Context / Metadata
   • Magnifies management confusion
   • Ramps up redundant work
   • Atrophies operational agility
   • Breeds bad business decisions
   • Complicates communication
   • Compounds compliance risk
   • Tramples data trust
   • Creates reports / analysis conflicts
These are just some of the more common issues that diminish the value of your data – you are dealing with at least a few of them now. Restoring data value requires data governance orchestration across these issues, and beyond.
In part, DG efforts can fail for the same reason any organization program fails...

- Failure to establish and communicate a compelling vision
- Poor planning
- Political naïveté
- Resistant culture
- Poorly defined objectives
- Lack of support (executive & otherwise)
- Poor ongoing communication and selling of the program
- Picked wrong success measures
- Unmanaged expectations
- Lack of leadership
- Etc.
...but there are additional common reasons DG programs fail

- “Overhead” perception
- Purely IT program positioning
- Raised visibility too early
- Failure to build organizational alliances
- Not positioned as a business enabler
- Leader too junior
- Run by a data person who is not a strong people person
- Velocity expectations
DG should optimize the value of data to the organization

To optimize data value we can lower data costs and/or increase data worth

**Lower Data-Related Costs**
- Reduce redundant data-related work
- Rationalize data applications
- Rationalize data vendor relationships
- Manage data retention
- Reduce risk-related costs

**Increase Data Worth**
- Create shared understanding of data
- Ensure data quality
- Ensure data timeliness
- Establish trust in data
- Make finding right data easier
In DG, consider the value of being able to answer these questions…

• Where can I find the information I need?
• What does this data mean?
• Is this data good enough for my needs?
• What am I allowed to do with this data?
• Who can help me if I have questions about this data?
Metadata Management & DG

Diminishing Value of Data in need of DG
- Lack of common names, definitions, and context
- Inadequate data accountability
- Inconsistent and undefined business and data quality rules
- Information security confusion
- Deficient data change control
- Ineffective communication
- Unidentified source of data

Metadata Management
- Business Glossary
- Business Process Glossary
- Business Rule Management
- Risk Management
- Rules and Regulations
- Data Lifecycle Management
- Data Ownership Management
- Data Quality Rule Management
- Data Security Management
- Impact Analysis
- Audit Trail
Metadata Management &

DATA STEWARDSHIP
What is Data Stewardship?

- Data stewardship encompasses the tactical management and oversight of the company’s data assets
- Data stewardship is generally a business function facilitating the collaboration between business and IT, and driving the correction of data issues
- Several models for Data Stewardship:
  - By Domain or Entity
  - By Business Function
  - By System
  - By Business Process
  - By Project
Metadata Management & Stewardship

Data Steward Responsibilities

Strategy & Planning
- Understand strategy as relates to data area
- Define & recommend data enhancement projects
- Provide feedback on SBL standards

Project Scoping & Analysis
- Help identify data sources
- Review & provide feedback on project
- Support development of test plan

Project Develop, Test & Deploy
- Define DQ needs & make sure is integrated into business requirements
- Help develop training on data

Daily Operations
- Monitor business changes for data impact
- Monitor DQ metrics and recommend corrective action, as needed
- Field questions about data area of responsibility

Metadata Management
- Business Definitions
- Business Process Definitions
- Business Rules
- Rules and Regulations
- Data Lifecycle, including lineage and transformations
- Data Quality Rules
- Data Security Rules
- Data Dictionary
- Context Definitions
- Impact Analysis
Metadata Management &

DATA QUALITY MANAGEMENT
What is Data Quality Management?

- Data Quality Management (DQM) is about employing processes, methods, and technologies to ensure the quality of the data meets specific business requirements.
- Trusted data delivered in a timely manner is the ultimate goal.
- DQM can be reactive or preventive. More mature companies are capable of anticipating data issues and prepare for them (that’s where Metadata Management is key).
- DQM encompasses many activities:
  - Data Profiling
  - Data Validation
  - Data Cleansing or Scrubbing
  - Data Consolidation
  - Data Matching
  - Data Survivorship
  - Data Standardization
  - Data Reconciliation
  - Data Enrichment
  - Data Monitoring
  - Data Quality Dashboards
  - Data Lineage and Traceability
  - Data Classification or Categorization
Datasource DQM Process

1. Profile
   - Data Steward
     - Person capable of running queries to determine characteristics of the data
     - Profiles:
       - Profiling Tool (e.g., Informatica Analyst)
       - SQL

2. Define Rules
   - Data Steward & Developer
     - Define validation rules
     - Define manipulation rules
     - Establish reference data
     - Data tool:
       - DQ Tool (e.g., Informatica Analyst/Developer)
       - MS Excel

3. Implement
   - Developer
     - Translate business rules into data quality reusable rules
     - Develop DQ objects (i.e., Scorecards)
     - Match / De-duplicate
     - DQ tool:
       - DQ Tool (e.g., Informatica Analyst/Developer, PowerCenter)

4. Monitor
   - Business collaborates with Data Steward/Developer
     - Re-profile
     - Measure quality
     - DQ Tool (e.g., Informatica Analyst)
     - BI Tool (e.g., Cognos, etc.)

5. Corrective Action
   - Data Steward & Developer
     - Update data in source, update reference data, OR update rule
     - Reconcile duplicates & trace survivors

Person who owns the data and capable of making changes in the system

Person capable of running queries to determine characteristics of the data
Typically, organizations approach Data Profiling as a 2D activity:
A Look into Data Profiling (2 of 3)

But Data Profiling is in a Spectrum:
A Look into Data Profiling (3 of 3)

Therefore, a 3\textsuperscript{rd} dimension must be considered, which is tied to metadata management:

![Diagram showing the relationship between Requirements, Data Profile Techniques, and Metadata Information. The Requirements point is connected to the Data Profile Techniques axis by a line, and the Metadata Information point is connected to the Requirements point by another line. The Data Profile Metric Results is shown in a separate chart.]
Besides Data Profiling, the other DQM activities can certainly benefit from Metadata Management:

- Data Validation
- Data Cleansing or Scrubbing
- Data Consolidation
- Data Matching
- Data Survivorship
- Data Standardization
- Data Reconciliation
- Data Enrichment
- Data Monitoring
- Data Quality Dashboards
- Data Lineage and Traceability
- Data Classification or Categorization
Collaborative Data Management

Metadata Management

Data Governance

Data Quality Management

Data Stewardship
A Customer’s Story – Conclusion

Metadata Management
- Business definitions
  - Accounting NCO
  - Operational NCO
- Business and DQ rules around NCO
- Sources of related data, attributes and their lineage/transformations
- NCO usage by different business processes and reports

Data Governance
- Business alignment
- Enterprise standards
- Risk mitigation
- Dispute resolution
- Impact analysis
- Ownership assignment

Data Quality Management
- Monitoring of NCO quality
- Dashboards and scorecards
- Alerts on suspicious accounts
- Reports on company performance and risk level

Data Stewardship
- Easy identification of sources/attributes related to NCO
- Proactive monitoring of thresholds and expected values
- Expedited issue resolution
- Streamlined maintenance
We are a consulting company that focuses exclusively on Enterprise Data Management & Business Intelligence, including both strategic and implementation services. We are passionate about data.

Datasource Consulting

STRATEGIC
Data Governance (People, Process, Tech)
Roadmaps & Assessments (BI, DW, EDM)
Program Management, Project Plans
Vendor Tool Selection

IMPLEMENTATION
Data Architecture, Data Integration
Data Quality, Business Intelligence
Master Data Management
Reporting & Analytics, Cloud

100% Success Rate | 80% of clients stay with Datasource over multiple years.

dcervo@datasourceconsulting.com