Applied A.I. Solutions

Foundations of Data Management

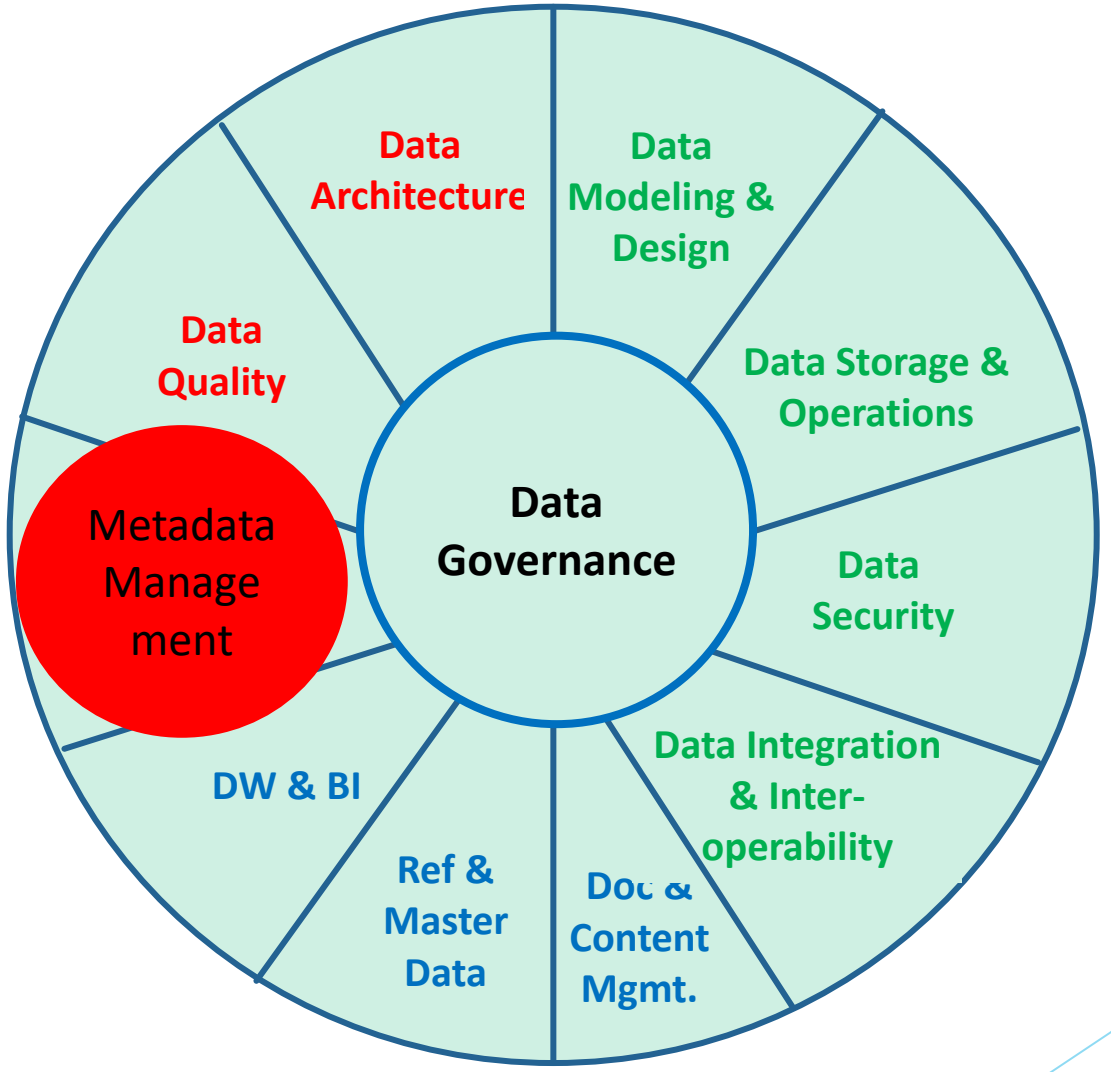
Professor

Daniel Vitaver-Bronstein, B.Sc., EMBA

daniel.vitaver-bronstein@georgebrown.ca

METADATA MANAGEMENT

The DAMA Wheel



[1] Main source: Copyright © 2017 DAMA International – DMBOK2 - Technics Publications, Basking Ridge, New Jersey, USA

1. INTRODUCTION

- Metadata is data about the asset (data)
- Metadata represents the context, information and knowledge about the data that we need to manage
- Metadata helps to interpret data in a meaningful way
- Metadata originates from a range of processes related to data creation, processing, and use, including architecture and governance

[1] Main source: Copyright © 2017 DAMA International – DMBOK2 - Technics Publications, Basking Ridge, New Jersey, USA

4

MM Framework

Definition

Planning, Implementation, and Control activities to enable access to high-quality, integrated metadata

[1] Main source: Copyright © 2017 DAMA International – DMBOK2 - Technics Publications, Basking Ridge, New Jersey, USA

Goals

1. Provide organizational understanding of business terms and lineage
2. Collect and integrate metadata from diverse sources
3. Provide a standard way to access metadata
4. Ensure Metadata Quality and Security

Business Drivers

[1] Main source: Copyright © 2017 DAMA International – DMBOK2 - Technics Publications, Basking Ridge, New Jersey, USA

MM programs are guided by the following principles:

- Organizational commitment
- Strategy
- Enterprise perspective
- Socialization
- Access
- Quality
- Audit
- Improvement

https://www.gartner.com/doc/reprints?id=1-24L1M1LE&ct=201112&st=sb

[1] Source: Copyright © 2017 DAMA International – DMBOK2 - Technics Publications, Basking Ridge, New Jersey, USA

Inputs

- Business requirements
- Metadata issues
- Data architecture
- Business metadata
- Technical metadata
- Operational metadata
- Data governance metadata

[1] Source: Copyright © 2017 DAMA International – DMBOK2 - Technics Publications, Basking Ridge, New Jersey, USA

Activities

1. Define metadata strategy
2. Understand metadata requirements
3. Define Metadata architecture
   - Create metamodel
   - Apply metadata standards
   - Manage metadata stores
4. Create and maintain metadata
   - Integrate metadata
   - Distribute and deliver metadata
5. Query, report and analyze metadata

[1] Source: Copyright © 2017 DAMA International – DMBOK2 - Technics Publications, Basking Ridge, New Jersey, USA

Deliverables

1. Metadata strategy
2. Metadata standards
3. Metadata architecture
4. Metamodel
5. Unified metadata
6. Data lineage
7. Impact analysis
8. Dependency analysis
9. Metadata control process

[1] Source: Copyright © 2017 DAMA International – DMBOK2 - Technics Publications, Basking Ridge, New Jersey, USA

Suppliers	Participants	Consumers
• Business data stewards • Data managers • Data governance bodies • Data modelers • Database administrators	• Data stewards • Project managers • Data architects • Business analysts • System analysts	• Application developer analyst • Data integrators • Business users • Knowledge workers • Customers, collaborators • Data scientists • Data journalists

[1] Source: Copyright © 2017 DAMA International – DMBOK2 - Technics Publications, Basking Ridge, New Jersey, USA

Technical Drivers

Techniques

- Data lineage and impact analysis
- Metadata for Big Data ingest

Tools

- Metadata repository management tools

Metrics

- Metadata coverage scorecard
- Metadata repository contribution
- Metadata usage reports
- Metadata quality scorecard

[1] Source: Copyright © 2017 DAMA International – DMBOK2 - Technics Publications, Basking Ridge, New Jersey, USA

Drivers

- Increase confidence in data
- Increase the value of strategic information
- Improve operational efficiency
- Improve communication between data consumers, IT professionals
- Prevent the use of out-to-date or incorrect data
- Create accurate impact analysis thus reducing risks
- Support regulatory compliance

[1] Source: Copyright © 2017 DAMA International – DMBOK2 - Technics Publications, Basking Ridge, New Jersey, USA

Essential Concepts

1. Metadata vs Data
2. Type of Metadata

Business Metadata	Technical Metadata	Operational Metadata
• Definitions, descriptions of data sets, tables and columns • Business, transformation rules, calculations • Data Models • DQ rules, measurements • Data update schedules	• Physical database table, column names, properties • Access permissions • Data CRUD • Physical data models • Models, assets relationships • ETL job details • File format schema definition	• Job execution logs (batch) • History of extracts • Schedule anomalies • Audit results • Error logs • Reports, query access, frequency, execution time

[1] Source: Copyright © 2017 DAMA International – DMBOK2 - Technics Publications, Basking Ridge, New Jersey, USA

3. Metadata Registry Standard (ISO/IEC 11179)

- Framework for the generation and standardization of data elements
- Basic attributes of data elements
- Rules and guidelines for the formulation of data definitions
- Naming and identification principles for data elements
- Registration of data elements

ISO (the International Organization for Standardization) and IEC (the International Electrotechnical Commission) https://www.iso.org/standard/60341.html

4. Metadata for Unstructured Data

- Descriptive
- Structural
- Administrative
- Bibliographic
- Record Keeping
- Preservation Metadata

[1] Source: Copyright © 2017 DAMA International – DMBOK2 - Technics Publications, Basking Ridge, New Jersey, USA

5. Sources of Metadata

- Application metadata repositories
- Business Glossary
- Business Intelligence Tools
- Configuration management Tools
- Data Dictionaries
- Data Integration Tools
- Database management and System Catalogs
- Data Mapping Management Tools
- Data Quality Tools
- Directories and Catalogs
- Event messaging Tools
- Modeling Tools and Repositories
- Reference Data Repositories
- Service Registries
- Other Metadata Stores

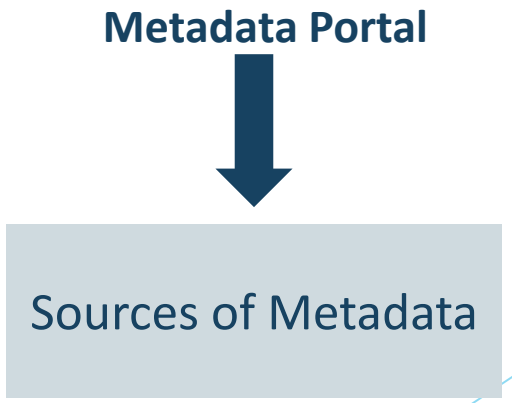
[1] Source: Copyright © 2017 DAMA International – DMBOK2 - Technics Publications, Basking Ridge, New Jersey, USA

- Distributed Metadata Architecture

  A single access point – (centralized)

  Pros:
  - Metadata always current, valid
  - Queries are distributed, improving performance
  - Simpler automated metadata query processing
  - No metadata replication or synchronization processes

  Cons:
  - Additions to Metadata are not supported
  - Limited standardization
  - Query limitations by source systems availability
  - Metadata quality depends on source systems



[1] Source: Copyright © 2017 DAMA International – DMBOK2 - Technics Publications, Basking Ridge, New Jersey, USA

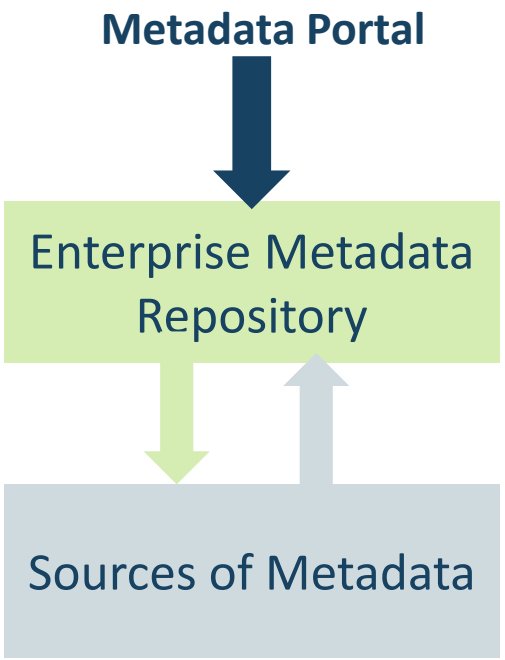
6. Metadata Architecture

- Centralized Metadata Architecture

  A single Metadata repository (distributed)

  Pros:
  - High availability
  - Quick Metadata retrieval
  - Improved quality database structures
  - Extracted Metadata enhancement

  Cons:
  - Complex processes
  - Costly Maintenance



[1] Source: Copyright © 2017 DAMA International – DMBOK2 - Technics Publications, Basking Ridge, New Jersey, USA

- Hybrid Metadata Architecture
  - Combines Centralized and Distributed architectures
  - A repository design only accounts for user-added Metadata, and critical standardized items

Pros:

- Near real-time retrieval of metadata from its source
- Enhanced metadata
- Reduced manual IT interventions and custom-coded functionality

Cons:

- Limited availability from source systems
- Increased overhead to get metadata results from the central repository

[1] Source: Copyright © 2017 DAMA International – DMBOK2 - Technics Publications, Basking Ridge, New Jersey, USA

2. ACTIVITIES

a) Define Metadata Strategy

- Initiate Metadata strategy planning
- Conduct key stakeholders interviews
- Assess existing metadata sources and information architecture
- Develop future Metadata architecture
- Developed a phased implementation plan

b) Understand metadata Requirements

- Volatility, synchronization, history, access rights, structure. integration, maintenance, management, quality, security

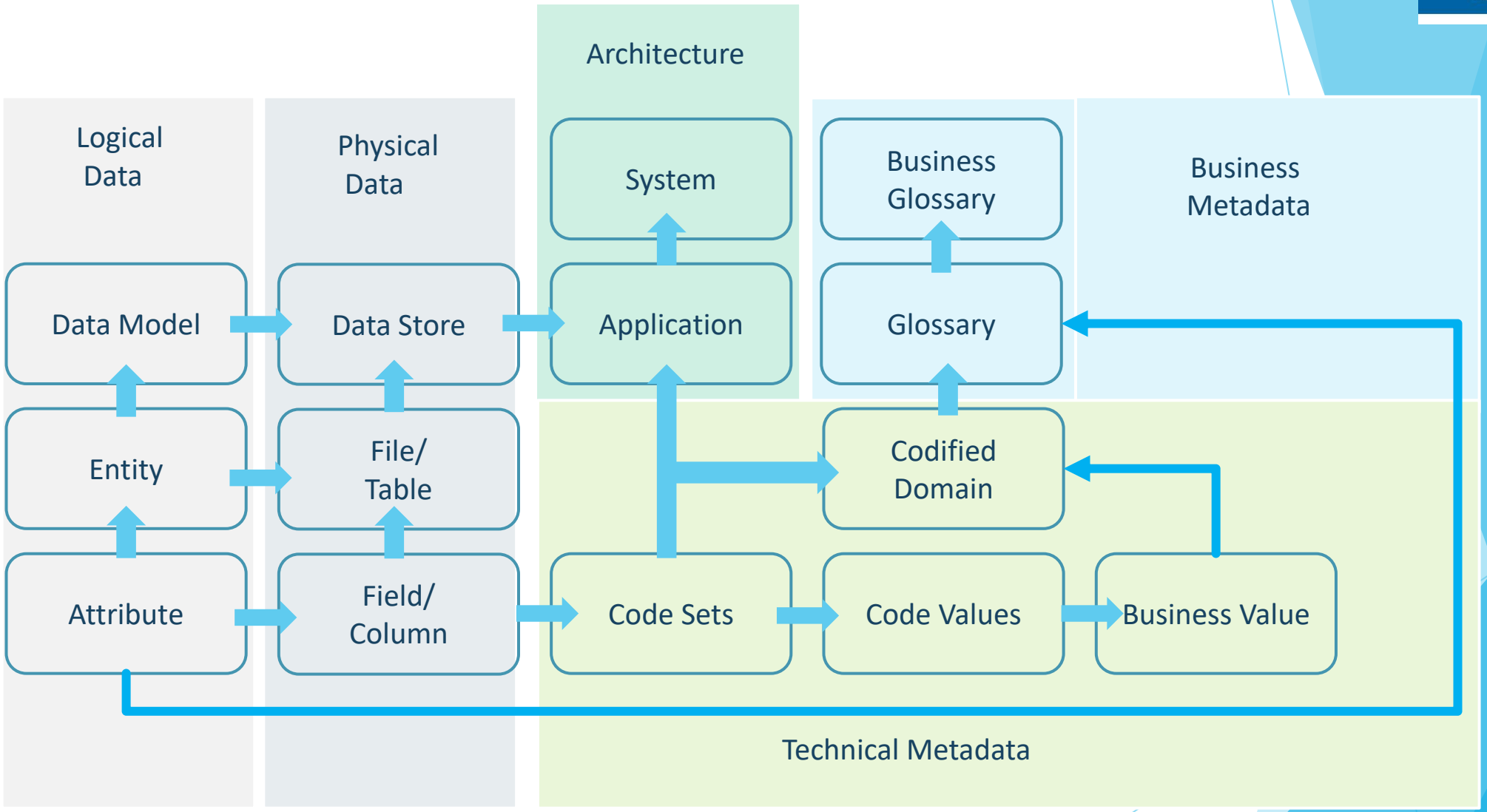
[1] Source: Copyright © 2017 DAMA International – DMBOK2 - Technics Publications, Basking Ridge, New Jersey, USA

c) Define Metadata Architecture

- Create a Metamodel (data model for metadata)
- Apply Metadata standards
- Manage Metadata Stores (repository environment)

[1] Source: Copyright © 2017 DAMA International – DMBOK2 - Technics Publications, Basking Ridge, New Jersey, USA

Example Metadata Repository Metamodel



[1] Source: Copyright © 2017 DAMA International – DMBOK2 - Technics Publications, Basking Ridge, New Jersey, USA

c) Define Metadata Architecture

- Create a Metamodel (data model for metadata)
- Apply Metadata standards
- Manage Metadata Stores (repository environment)

d) Create and Maintain Metadata

- Integrate
- Distribute and Deliver Metadata

e) Query, Report, and Analyze Metadata

[1] Source: Copyright © 2017 DAMA International – DMBOK2 - Technics Publications, Basking Ridge, New Jersey, USA

3. TOOLS

- Metadata Repository management Tools

4. TECHNIQUES

- Data Lineage and Impact Analysis
  - Implemented Lineage
  - Designed Lineage
  - Lineage discovery
    - Business Focus
    - Technical focus
- Metadata for Big Data Ingest

Magic Quadrant

Figure 1. Magic Quadrant for Metadata Management Solutions

Source: Gartner (November 2020)

[1] Source: Copyright © 2017 DAMA International – DMBOK2 - Technics Publications, Basking Ridge, New Jersey, USA

5. IMPLEMENTATION GUIDELINES

- Readiness Assessment / Risk Assessment
  - Error in judgement
    - Incorrect, incomplete or invalid assumptions
    - Lack of knowledge about the context of the data
  - Exposure to sensitive data
  - SMEs retention

- Organizational and Cultural Change
  - Senior Management Commitment

[1] Source: Copyright © 2017 DAMA International – DMBOK2 - Technics Publications, Basking Ridge, New Jersey, USA

6. METADATA GOVERNANCE

- Process Controls
- Documentation of Metadata Solutions
- Metadata Standards and Guidelines
- Metrics
  - Metadata Management Maturity
  - Metadata repository availability and completeness
  - Metadata Usage and Business Glossary Activity
  - Metadata documentation quality

[1] Source: Copyright © 2017 DAMA International – DMBOK2 - Technics Publications, Basking Ridge, New Jersey, USA