Applied A.I. Solutions

Foundations of Data Management

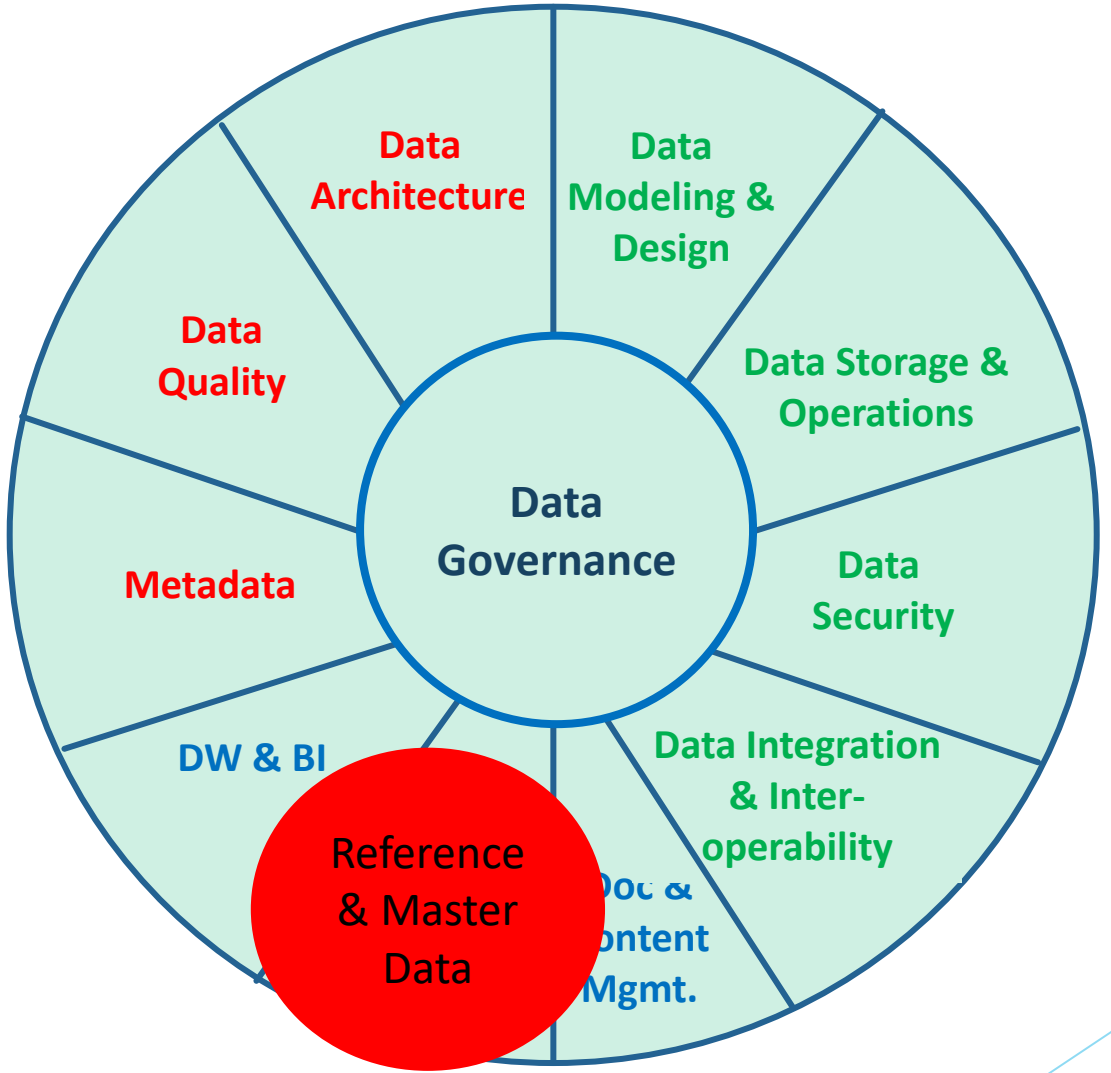
Professor

Daniel Vitaver-Bronstein, B.Sc., EMBA

daniel.vitaver-bronstein@georgebrown.ca

REFERENCE AND MASTER DATA

The DAMA Wheel



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

1. INTRODUCTION

- In any organization certain data is required across business areas, processes, and systems
- Everyone benefits is they can access to the same customer lists, geographic location codes, accounting costs centres, and other data used to run the business
- Reference and Master Data are used to avoid inconsistencies in data structures and data values between systems

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

R&MD Management Framework

Definition

Managing shared data to meet organizational goals, reduce risks associated with data redundancy, ensure higher quality, and reduce costs of data integration

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

Goals

1. Enable sharing of information assets across business domains and applications
2. Provide authoritative source of reconciled and quality-assessed master and reference data
3. Lower cost and complexity through use of standards, common data models, and integration patterns

Business Drivers

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

R&MD programs are guided by the following principles:

- Shared Data
- Ownership
- Quality
- Stewardship
- Controlled Change
- Authority

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

Inputs

- Business Drivers
- Cross functional requirements
- Industry standards
- Data glossary
- Purchased data and/or open data and code sets
- Business rules

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

Deliverables

1. Master and Reference Data requirements
2. Data Models and Integration patterns
3. Reliable Reference and Master Data
4. Reusable Data Services

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

Suppliers	Participants	Consumers
• SME • Data Stewards • Application developers • Data providers • Business analysts • Infrastructure systems analysts	• Data Architects • Data Quality Analysts • Data Modelers • Data Stewards • Data Analysts • Data Integrators	• Master Data analysts • Data Integrators • Data architects • Application users • Application developers • Solution architects

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

Techniques

- Conditions-of-use agreements
- Business key cross references
- Processing log analysis

Tools

- Data Modeling tools
- Metadata repositories
- Data profiling and quality tools
- MDM application platforms
- Data sharing / Integration architecture

Metrics

- DQ and compliance
- Data Change activity
- Data Consumption, services
- Data sharing availability
- Data Steward coverage
- Data sharing volume, utilization

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

Drivers

For Master Data Management

- Meeting organizational data requirements
- Managing data quality
- Managing the costs of data integration
- Reducing risk

For Reference Data Management

- Meeting data requirements for multiple initiatives
- Reduce costs of data integration
- Manage the quality of reference data

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

Essential Concepts

A. Master Data

- It is “the **data that provides the context for business activity** in the form of common and abstract concepts. It includes definitions and identifiers, and external objects involved in business transactions, such as **customers, products, employees, vendors, and other controlled domains** (code values)”

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

[2] Malcom Chisholm proposed a six-layer taxonomy of data that includes Metadata, Reference Data, enterprise structure data, transaction structure data, transaction activity data, and transaction audit data (Chisholm, 2008; Talburt and Zhou, 2015)

Master Data represents the business objects which contain the most valuable, agreed upon information shared across an organization

- A change in master data is always managed as a part of existing business processes
- Data of customers, products, employees, materials, suppliers and vendors are some examples of Master Data
- Systems of Record, System of Reference
- **Trusted Source**

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

https://pediaa.com/

Master Data Management

"A technology-enabled discipline in which **business and IT work together** to ensure the uniformity, accuracy, stewardship, semantic consistency, and accountability of the enterprise's official shared Master Data assets. " (Gartner)

- Planning for MDM includes several basic steps:
  - Identify master Data entities
  - Develop rules for matching and merging
  - Correct mismatches
  - **Distribute trusted data**

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

- Master Data Management Key Processing Steps
  - Data Model Management
  - Data Acquisition
  - Data Validation, Standardization, and Enrichment
  - Entity Resolution and Identifier Management
    - Matching
    - Identity Resolution
    - Matching Workflows / Reconciliation Types
    - Master Data ID Management
    - Affiliation Management
  - Data Sharing and Stewardship

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

- Customer/Client Master Data (CRM)
- Financial Master Data (ERP)
- Legal Master Data
- Product Master Data (MRP, ERP)
- Location Master Data (GIS)
- Industry Master Data – Reference Directories

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

B. Reference Data

"It is any data used to characterize or **classify other data**, or to relate data to information external to an organization (Chisholm, 2001)"

- Reference Data Structure (Lists, Taxonomies, Ontologies)
- Proprietary or Internal Reference Data (to support internal process and applications)
- Industry Reference Data
- Geographic or Geo-statistical Data
- Computational Reference Data
- Standard Reference Data Set Metadata

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

Reference Data is data that defines the set of **permissible values** to be used by other data fields

- A change in the reference data values may require an associated change in the business process to support the change

- **Units of measurement, country codes, corporate codes, fixed conversion rates, and calendar structure are some examples**

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

https://pediaa.com/

C. Data Sharing Architecture

- Registry
- Transaction Hub
- Hybrid of Registry and Transaction Hub (consolidated approach)

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

Activities

1. Identify drivers and requirements
2. Evaluate and assess data sources
3. Define architectural approach
4. Model data
5. Define stewardship and maintenance process
6. Establish governance policies
7. Implement data sharing / Integration services

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

2. ACTIVITIES

- Define Drivers and Requirements
- Evaluate and Assess Data Sources
- Define **Architectural** Approach
- **Model** Master Data or Reference Data Sets
- Define Stewardship and Maintenance Processes
- Establish Governance Polices to **Enforce** Use of Master Data or Establish Reference Data Governance Policies

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

2. TOOLS AND TECHNIQUES

3. IMPLEMENTATION GUIDELINES

- Adhere to Master Data Architecture
- Monitor Data Movement
- Manage Reference Data Change
- Data Sharing Agreements

4. ORGANIZATION AND CULTURAL CHANGE

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

REFERENCE AND MASTER DATA GOVERNANCE

6. Governance processes will determine:

- Data sources to be integrated
- Data quality rules to be enforced
- Conditions of use rules to be followed
- Activities to be monitored and the frequency of monitoring
- Priority and response level of data stewardships efforts
- How information is to be represented to meet stakeholder needs
- Approval Standards, R&MD deployment expectations

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

- **Metrics**
  1. Data Quality and Compliance
  2. Data Change Activity
  3. Data Ingestion and Consumption
  4. SLA level of adherence
  5. Data Steward Coverage
  6. Total Cost of Ownership
  7. Data Sharing Volume and Usage

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