

DATA STEWARDSHIP BODY OF KNOWLEDGE (DSBOK)

Release 2.2 August 2013.

This document was created in collaboration of the leading experts and educators in the field and members of the Certified Data Steward (CDS) Advisory Council.



1. Data Stewardship Fundamentals

1.1. Definitions

- 1.1.1. Stewardship
- 1.1.2. Data Stewardship

1.2. Data Stewardship Organizations

- 1.2.1. Kinds of Data Stewards
 - 1.2.1.1. Kinds of Business Stewards
 - 1.2.1.2. Kinds of IT Stewards
 - 1.2.1.3. Differences of Business and IT Stewards
- 1.2.2. Roles and Responsibilities
- 1.2.3. Reporting and Relationships

1.3. Data Steward Characteristics

- 1.3.1. Knowledge and Experience
- 1.3.2. Skills and Competencies
 - 1.3.2.1. Data Skills
 - 1.3.2.2. Technical Skills
 - 1.3.2.3. Human Skills
 - 1.3.2.4. Facilitation Skills
 - 1.3.2.5. Communication Skills

2. Data Management Processes

2.1. Architectural Processes

- 2.1.1. Enterprise and Subject Area Modeling
- 2.1.2. Data Mapping and Consolidation
- 2.1.3. Data Flow

2.2. Utilization Processes (CRUD)

- 2.2.1. Create Processes and Practices
- 2.2.2. Retrieval and Reporting Processes and Practices
- 2.2.3. Update Processes and Practices
- 2.2.4. Delete and Archive Processes and Practices
- 2.2.5. Business Process Management

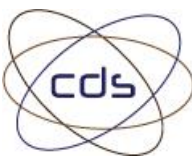
2.3. Custodial Processes

- 2.3.1. Database Administration
- 2.3.2. Security Administration and Access Authorization
- 2.3.3. Backup, Recovery, and Business Continuity

3. Information Management Concepts

3.1. Types of Data and Information

- 3.1.1. Operational and Analytical
- 3.1.2. Event and Reference
- 3.1.3. Structured and Unstructured



- 3.1.4. Transactional Data
- 3.1.5. Master Data

3.2. *Types of Data Stores*

- 3.2.1. Application Databases
- 3.2.2. Departmental and End-User Databases
- 3.2.3. Data Warehouses and Data Marts
- 3.2.4. Operational Data Stores
- 3.2.5. Master Data Hubs

3.3. *Common Uses of Data*

- 3.3.1. Record Keeping and Audit Trail
- 3.3.2. Reporting and Information
- 3.3.3. Measurement and Monitoring
- 3.3.4. Analysis and Discovery

3.4. *Business Data Flow*

- 3.4.1. The Business Value of Data
- 3.4.2. Data Sources and Data Acquisition
- 3.4.3. Data Providers and Consumers
- 3.4.4. Data Flow through Organizations and Systems
- 3.4.5. Data Conversion and Consolidation
- 3.4.6. Data Replication and Redundancy
- 3.4.7. Data Sharing and Interfaces
- 3.4.8. Data Disposal and Destruction

3.5. *Information Management Disciplines*

- 3.5.1. Data Modeling
- 3.5.2. Metadata Management
- 3.5.3. Content Management
- 3.5.4. Enterprise Information Management
- 3.5.5. Data Quality
- 3.5.6. Data Governance
- 3.5.7. Data Integration
- 3.5.8. Data Warehousing
- 3.5.9. Master Data Management
- 3.5.10. Business Intelligence
- 3.5.11. Business Analytics
- 3.5.12. Performance Management
- 3.5.13. Data Mining
- 3.5.14. Predictive Analytics

4. *Data Quality*

4.1. *Quality Management Basics*

- 4.1.1. Quality Perspectives
 - 4.1.1.1. Expectations
 - 4.1.1.2. Purpose
 - 4.1.1.3. Specifications



- 4.1.1.4. Defects
- 4.1.2. Quality Management Terminology
 - 4.1.2.1. Quality Control (QC)
 - 4.1.2.2. Quality Assurance (QA)
 - 4.1.2.3. Waste and rework
 - 4.1.2.4. Inspection, Correction, Prevention
- 4.1.3. Quality Management Methodologies
 - 4.1.3.1. Total Quality Management (TQM)
 - 4.1.3.2. Statistical Process Control (SPC)
 - 4.1.3.3. Six Sigma
- 4.2. *Data Quality Concepts and Principles*
 - 4.2.1. Data Quality Definitions
 - 4.2.1.1. Defect Free
 - 4.2.1.2. Conform to Specifications
 - 4.2.1.3. Fit to Purpose
 - 4.2.1.4. Meet Customer Expectations
 - 4.2.2. Common Causes of Data Quality Problems
 - 4.2.3. Costs and Benefits of Data Quality
- 4.3. *Data Quality Dimensions*
 - 4.3.1. Content Quality (Correctness)
 - 4.3.1.1. Accuracy
 - 4.3.1.2. Completeness
 - 4.3.1.3. Precision
 - 4.3.1.4. Granularity
 - 4.3.1.5. Consistency
 - 4.3.2. Structural Quality (Integrity)
 - 4.3.2.1. Identity
 - 4.3.2.2. Reference
 - 4.3.2.3. Cardinality
 - 4.3.2.4. Dependency
 - 4.3.2.5. Inheritance
 - 4.3.2.6. Domain of Values
 - 4.3.3. Temporal Quality (Timeliness)
 - 4.3.3.1. Currency
 - 4.3.3.2. Retention
 - 4.3.3.3. Continuity
 - 4.3.3.4. Precedence
 - 4.3.4. Business Quality (Value)
 - 4.3.4.1. Aligned
 - 4.3.4.2. Trusted
 - 4.3.4.3. Understandable
 - 4.3.4.4. Reliable
 - 4.3.4.5. Compliant
 - 4.3.4.6. Useful
 - 4.3.5. Usage Quality (Usability)
 - 4.3.5.1. Available
 - 4.3.5.2. Accessible
 - 4.3.5.3. Navigable
 - 4.3.5.4. Recoverable



- 4.3.5.5. Secure
- 4.3.5.6. Private
- 4.3.6. Presentation Quality (Communication)
 - 4.3.6.1. Clear
 - 4.3.6.2. Organized
 - 4.3.6.3. Non-Ambiguous
 - 4.3.6.4. Tool-Integrated
 - 4.3.6.5. Media-Fit

4.4. *Data Quality Processes and Projects*

- 4.4.1. Data Quality Assessment
- 4.4.2. Root Cause Analysis
- 4.4.3. Data Cleansing
- 4.4.4. Data Enrichment

4.5. *Data Quality in IT Processes and Projects*

- 4.5.1. Data Quality in Application and Database Projects
- 4.5.2. Data Quality in Data Conversion and Consolidation Projects
- 4.5.3. Data Quality in Data Sharing and Data Interfaces

5. **Data Governance**

5.1. *Data Governance Basics*

- 5.1.1. Definitions
- 5.1.2. Purpose
- 5.1.3. What to Govern and What not to Govern?

5.2. *Components of Data Governance*

- 5.2.1. Governance Goals
 - 5.2.1.1. Quality
 - 5.2.1.2. Security
 - 5.2.1.3. Compliance
 - 5.2.1.4. Standardization
 - 5.2.1.5. Usage
 - 5.2.1.6. Value
 - 5.2.1.7. Business Impact
- 5.2.2. Governance Elements
 - 5.2.2.1. Roles
 - 5.2.2.2. Responsibilities
 - 5.2.2.3. Decision Authority
 - 5.2.2.4. Accountability
- 5.2.3. Policies
 - 5.2.3.1. Quality
 - 5.2.3.2. Sensitivity
 - 5.2.3.3. Privacy
 - 5.2.3.4. Security
 - 5.2.3.5. Access
 - 5.2.3.6. Retention
 - 5.2.3.7. Disposal



- 5.2.4. Standards
 - 5.2.4.1. Naming
 - 5.2.4.2. Definition
 - 5.2.4.3. Architecture
 - 5.2.4.4. Transfer
 - 5.2.4.5. Integration
- 5.2.5. Technology
 - 5.2.5.1. Data Governance Enabling
 - 5.2.5.2. Data Quality
 - 5.2.5.3. Workflow
 - 5.2.5.4. Collaboration

5.3. *Data Governance Programs*

- 5.3.1. Business Drivers for Governance
 - 5.3.1.1. Legal
 - 5.3.1.2. Regulatory
 - 5.3.1.3. Financial
 - 5.3.1.4. Operational
 - 5.3.1.5. Competitive
- 5.3.2. Data Governance Roles
 - 5.3.2.1. Sponsors
 - 5.3.2.2. Owners
 - 5.3.2.3. Data Stewards
 - 5.3.2.4. Custodians
 - 5.3.2.5. Stakeholders
 - 5.3.2.6. Data Officers
 - 5.3.2.7. Councils, Committees, Competency Centers

5.4. *Executing Data Governance*

- 5.4.1. Getting Started
- 5.4.2. Funding and Sponsorship
- 5.4.3. Day-to-Day Governance
- 5.4.4. Sustaining Governance
- 5.4.5. Monitoring Governance: Measures, Metrics, Maturity
- 5.4.6. Evolving Governance
- 5.4.7. Stakeholder Communications

6. *Metadata Management*

6.1. *Metadata Concepts*

- 6.1.1. Metadata Definition
- 6.1.2. Roles of Metadata
 - 6.1.2.1. Classify
 - 6.1.2.2. Describe
 - 6.1.2.3. Guide
 - 6.1.2.4. Control
- 6.1.3. Types of Metadata
 - 6.1.3.1. Business Metadata
 - 6.1.3.2. Technical Metadata



6.1.4. Metadata Standards and Practices

6.2. *Data Modeling*

6.2.1. Data Model Principles

6.2.2. Types of Data Models

6.2.2.1. ER vs. Dimensional

6.2.2.2. Logical vs. Physical

6.2.2.3. Who and Why for each Type of Model

6.2.3. Complementary Models

6.2.3.1. Subject Area Models

6.2.3.2. State Transition Models

6.2.4. Reading Data Models

6.3. *Data Profiling*

6.3.1. Roles of Profiling

6.3.1.1. Support of Data Quality Management

6.3.1.2. Support of Master Data Management

6.3.1.3. Support of Data Migration

6.3.1.4. Support of Data Integration

6.3.2. Profiling Techniques

6.3.2.1. Column Profiling

6.3.2.2. Profiling Data Models

6.3.2.3. Profiling Time-Dependent Data

6.3.2.4. Subject Profiling

6.3.2.5. Profiling State-Transition Models

6.3.2.6. Attribute Dependency Profiling

6.3.2.7. Dynamic Data Profiling

6.3.3. Profiling Challenges

7. *Master Data Management*

7.1. *Master Data Management (MDM) Basics*

7.1.1. Master Data Definition

7.1.2. Sources of Master Data

7.1.3. Types of Master Data

7.1.4. MDM Objectives

7.1.5. MDM Architectures and Styles

7.1.6. Costs and Benefits of MDM

7.1.7. MDM Challenges and Best Practices

7.2. *Data Parsing, Matching, and De-Duplication*

7.2.1. Data Parsing and Standardization

7.2.1.1. Personal and Business Name

7.2.1.2. Postal Addresses

7.2.1.3. Geocoding

7.2.1.4. Date and Time

7.2.2. Data Matching and De-Duplication

7.2.2.1. Deterministic vs. Probabilistic

7.2.2.2. Matching Rules



- 7.2.2.3. Data Survivorship
- 7.2.2.4. Types of Matching Tools

7.3. *External Reference Data*

- 7.3.1. Address Directories
- 7.3.2. Business Directories
- 7.3.3. DUNS and EPC

7.4. *Global Data*

- 7.4.1. Format Differences across Countries and Regions
 - 7.4.1.1. Address Formats
 - 7.4.1.2. Personal Names
 - 7.4.1.3. Business Names
 - 7.4.1.4. Phone Numbers
 - 7.4.1.5. Job Titles
- 7.4.2. Special Characters and Diacritics
 - 7.4.2.1. Regional Alphabets
 - 7.4.2.2. Character Sets
 - 7.4.2.3. Code Pages and Unicode
- 7.4.3. Cultural and Legal Issues
 - 7.4.3.1. Privacy Laws
 - 7.4.3.2. Cultural Differences
- 7.4.4. Best Practices of Working with Global Data

Note: While this document goes into four levels of detail, the fourth level is not exhaustive and in many cases only includes some examples for topic clarification.