



ARE YOU READY TO ACE THE **CDMP EXAM**

Imagine the achievement

Imagine having a certification recognized in 90+ countries proving your data expertise.

It's not impossible

The CDMP looks tough, but with the right strategy and cheatsheets, it's achievable.

Cheat Sheets in action

We'll solve sample questions together and see how cheatsheets make it simple.

Clear promise

By the end, you'll have practical guides and a clear plan to pass on your first try.

[edward-gavier-garzon](#)

(+57) 3213352615



CHAPTER 1: DATA MANAGEMENT

Definition: Data Management is the development, execution and supervision of plans, policies, programs and practices that deliver, control, protect and enhance the value of data and information assets throughout their lifecycles.

Goals

- Understand and support Data Needs
- Capture, Store, Protect
- Ensure Data Quality
- Ensure Privacy
- Prevent Unauthorised Use
- Ensure Effective Use of Data

Essential Concepts

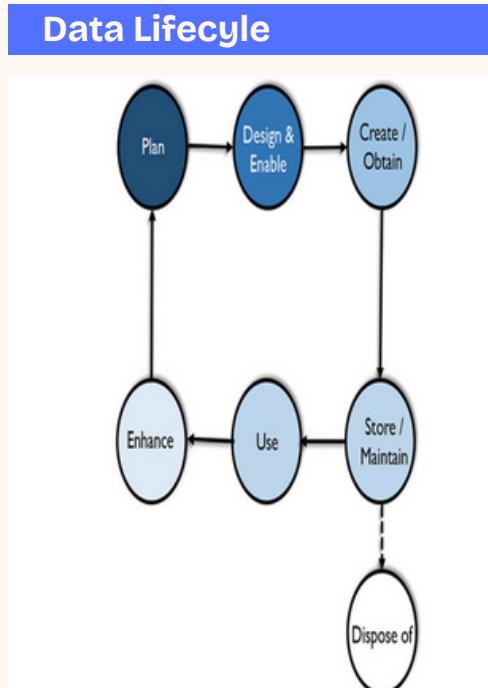
Data	DM Principles	Strategy
Information Metadata Asset	Leadership (CDO, Committed, Strategic) Valuable, Business Requirements Cross Functional, Lifecycle	DM Charter DM Scope Statement DM Implementation Roadmap

Challenges

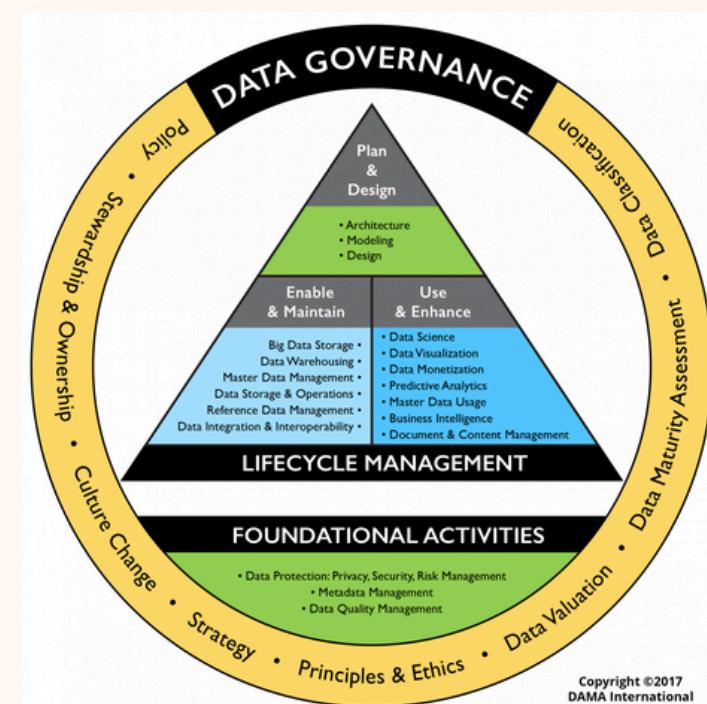
Differs from other Assets - Data Valuation - Data Quality - Planning (IT - Business) - Metadata
Enterprise Perspective - Lifecycle - Different Types of Data (Type, Content, Format, Protection)
Risks - Leadership Commitment

Business Drivers

- Competitive Advantage
- Manage Data as Capital
- Get Value from Data Assets



DAMA Wheel Evolved



CHAPTER 2: DATA HANDLING ETHICS

Definition: Data Handling Ethics are concerned with how to procure, store, manage, interpret, analyse/apply and dispose of data in ways that are aligned with ethical principles, including community responsibility.



Goals

- To define ethical handling of data in the organisation
- To educate staff on the organisational risks of improper data handling
- To change / instill preferred culture and behaviours on handling data
- To monitor regulatory environment, measure, monitor and adjust organisation approaches for ethics in data



Legal Changes
Employees Aware of Obligations
Oversight (Standards and Policies)
Review Plans



Review existing practices
Identify Risks
Adopt socially responsible risk model
Practice Gaps
Staff (Communicate- Educate)
Monitor and Maintain



Essential Concepts



Core

Impact on people
Potential for misuse
Value of data



Risks

Public key

- Spyware
- Trojan Horse
- Virus
- Worm



Principles

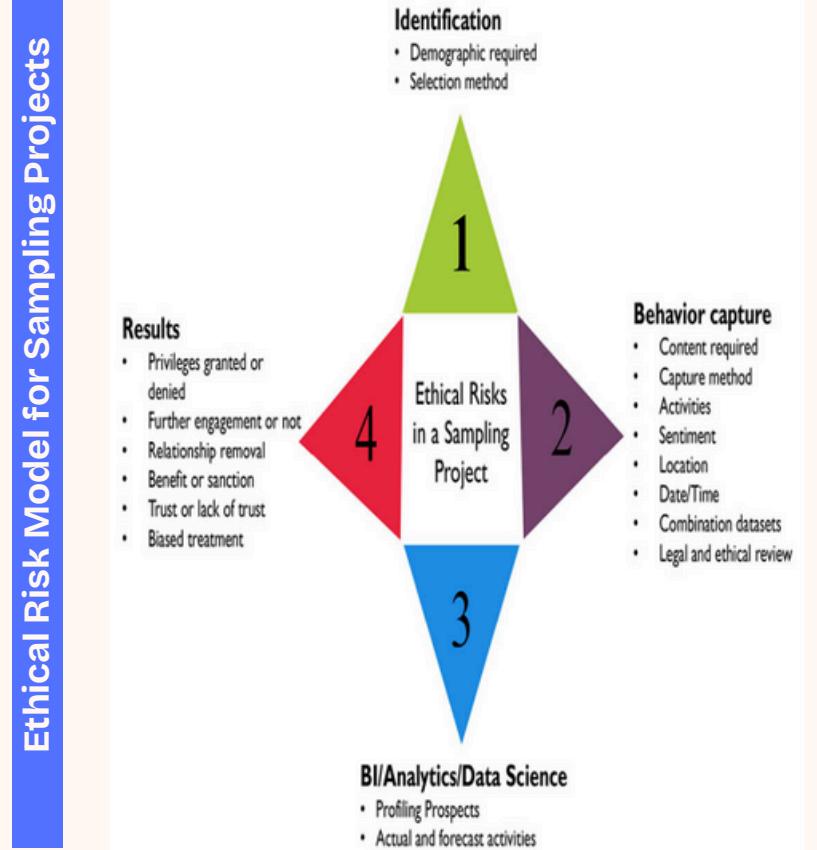
Ethical

- Respect for persons
- Beneficence
- Non-maleficence
- Value of data

Data Privacy Law

- Transparency
- Firness/Lawfulness
- Purpose
- Minimisation

- Accuracy
- Storage Limitation
- Integrity and Confidentiality
- Accountability



CHAPTER 3: DATA GOVERNANCE

Definition: The exercise of authority, control and shared decision making (planning, monitoring and enforcement) over the management of data assets

Goals

- Enable an organisation its data as an asset
- Define, approve, communicate and implement principles, policies, procedures, metrics, tools and responsibilities of data management
- Monitor and guide policy compliance, data usage and management activities

Governance

- Metrics
- Value
 - Effectiveness
 - Sustainability

Implementation

Organisation and Culture Adjustment and Communication

Technology

- Online Presence
- Business Glossary
- Workflow Tools
- Document and management
- DG Scorecards

Essential Concepts

DGO	Data Stewards	Activities	Types	Principles
CDO DG Steering Committee DG Council DGC DG Office DGC DG Office DGO Data Stewardship Teams Local DG Committees		<ul style="list-style-type: none">MetadataRules/StandarsDQ issuesOperational DG Issues	<ul style="list-style-type: none">Chief DSExecutive DSEnterprise DSBusiness DSData OwnerTechincal DSCo-ordinating DS	Enterprise Data must be <ul style="list-style-type: none">ModelledMaintened (Source-DQ)SecuredAccesableMetadata (Recorded-Utilised)

Data Policies

Data Policies	Activities
"What" Few Directives	<p>Define DG Readiness Assessment</p> <ul style="list-style-type: none">AvailabilityExecutionService LevelsEnvironments <p>Discovery</p> <p>Organisation Touch points</p> <p>DG Strategy</p> <ul style="list-style-type: none">CharterOperating FrameworkAccountabilitiesImplementation roadmapPlan for sucess <p>DG Operating Framework</p> <p>Develop</p> <ul style="list-style-type: none">GoalsPrinciplesPolices

Data Asset

Data Asset	DG & DM	Execution
Accountability Asset Audit Due Diligence Going Concern Level of validation Liability Quality Risk Value		<p>Change management</p> <p>Issue Management</p> <p>Regulatory compliance</p> <p>Implement DG</p> <p>Standars and Procedures (Sponsor)</p> <p>Business Glossary (Develop)</p> <p>Data Asset Valuation (Sponsor)</p> <p>Embed DG</p>



CHAPTER 4: DATA ARCHITECTURE

Definition: Identifying the data needs of the enterprise (regardless of structure), and designing and maintaining the master blueprints to guide data integration, control data assets, and align data investments with business strategy

Goals

- Identify data storage and processing requirements
- Design structures and plans to meet the current and long-term data requirements of the enterprise.
- Strategically prepare organisations to quickly evolve their products, services and data to take advantage of business opportunities in tech

Governance

- Oversight
- Manage (Designs, Lifestyle, Tools)
- Standards
- Artifacts

Implementation

- Readiness / Risk Assessment
- Organisations & Culture Change

Technology

- | | |
|--|----------------------------|
| Tools | Techniques |
| <ul style="list-style-type: none"> Modeling Asset Management | Lifestyle Projections |
| Graphic Design | Future, Active, Retirement |

Essential Concepts

EA Domains

- Business
- Data
- Applications
- Technology

EDA

- EDM: CDM-LDM
- Subject Areas
 - Bottom up
 - Top Down
 - Data Flow Design

EA Frameworks

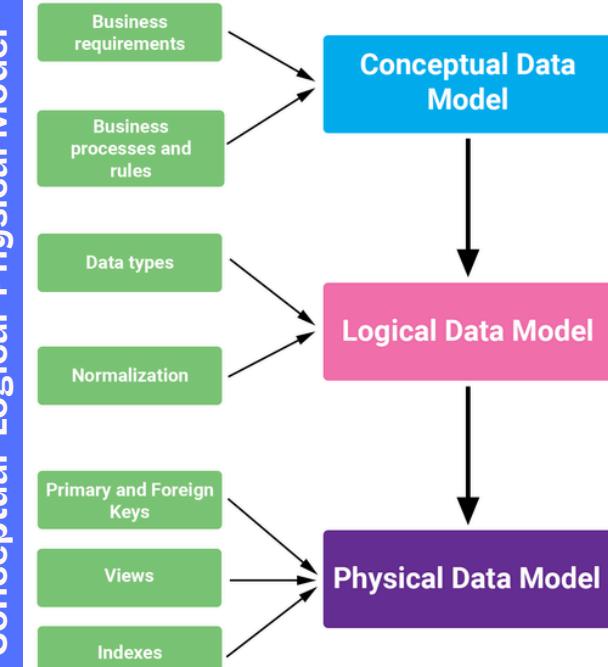
Zachman

- Columns (5W&H)
- Rows (Perspectives)

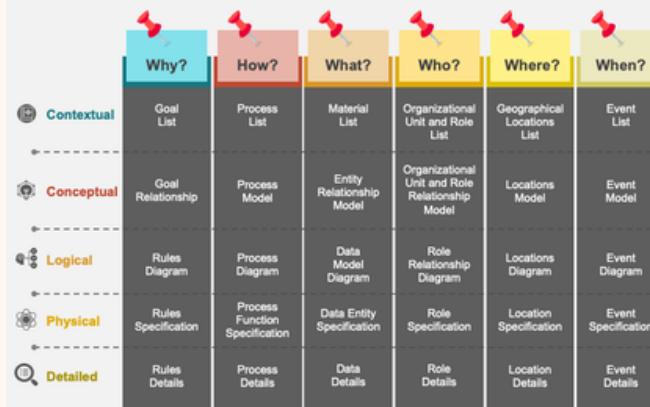
Activities

- Establish DA Practice
 - Strategy
 - Culture
 - Organisation
 - Working Methods
 - Results
- Evaluate Existing DA Roadmap
- Manage Requirements within Projects
 - Scope
 - Business Requirements
 - Design
 - Implement
 - Methodologies:
 - Waterfall
 - Incremental
 - Agile/Iterative
- Integrate with EA

Conceptual Logical Physical Model



ZACHMAN FRAMEWORK



CHAPTER 5: DATA MODELING AND DESIGN

Definition: Data modeling is the process of analysing and scoping data requirements, and then representing and communicating these data requirements in a precise form called a data model.



Goals

- To confirm and document an understanding of different perspectives, which leads to applications that more closely align with current and future business requirements, and creates a foundation to successfully complete broad-scoped initiatives such as master data management and data governance programs



Business Drives

- Common Vocabulary
- Document Knowledge
- Communication



Governance

Design Quality

Develop Standards: Data Modeling and Design

Review: Data Model Quality and Database Design

Manage: Data Model Versions and Integration

Metrics: Data Model Scorecard



Implementation



Technology

Tools

- Data Modeling
- Data Lineage
- Data Profiling

Techniques

- Best Practices
- Naming Standards
- Database design



Essential Concepts

Components

- | | |
|---|---|
| Entities | Relationships |
| <ul style="list-style-type: none">Thing (Noun)5W+HDefinitions | <ul style="list-style-type: none">VerbCardinalityArityForeign Keys |

Attributes

- Columns
- Domains
- Keys
 - Construction: Simple, surrogate, compound, composite
 - Function: Candidate, primary, alternate, foreign
- Domains
 - Constrains-Rules
 - Data Type
 - Format
 - List of Values
 - Range
 - Rule-based

Scheme

Relational

- Captures Business Rules
- Operational Systems
- No Redundancy

Dimensional

- Analysis
- Answers Business Questions
- Tables: Fact and Dimensions

Levels of Detail

Conceptual CDMDS

- High-Level Data
- Business Entities
- Relationships (Business Rules)

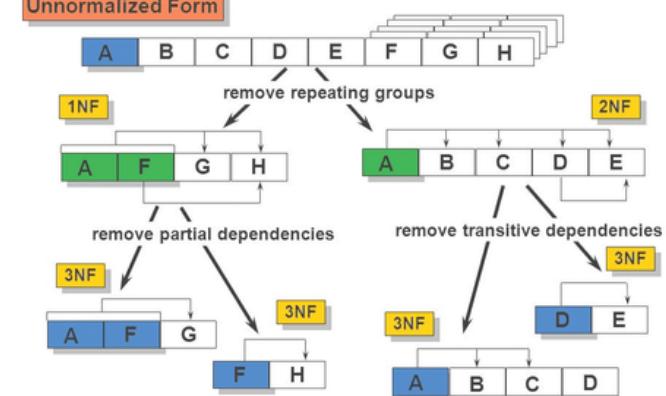
Logical LDM

- Attributes
- Normalisation
- Keys

Physical PDM

- Technical Solution
- DBMS

Unnormalized Form



Normalisation

1NF

- Valid PK, Atomic, Depend on key

2 NF

- Minimal PK, Depend Completely

3 NF

- No hidden PKs, no attributes

BCNF

- Hidden Business Rules

4 NF - 5 NF

Abstraction

- Generalisation - Specialisation



CHAPTER 6: DATA STORAGE AND OPERATIONS

Definition: The design, implementation and support of stored data to maximise its value.



Goals

- Manage availability of data throughout the data lifecycle
- Ensure integrity of data assets
- Manage performance of data transactions



Essential Concepts

Terms

Database
Instance
Schema
Node
Database Admon.

Administrators

Production DBA
Application DBA
Procedural DBA
Development DBA
NSA

Database Architectures

Centralised (One location)
Distributed (Replicated)
Federated (Coupling-Loose-Tight)
Blockchain
Cloud (VW, DaaS, Hosting)



Business Drives

Business Continuity
Principles

- Automation - Reuse - Best Practices
- Standars to SLAa - Projects



Governance

Metrics: Storage - Performance - Operational - Service

Asset Tracking: Software Licensing - Support

Data Audits - Data Validation



Implementation

Readiness / Risk Assessment
Organisation and Cultural Change



Technology

- Tools
- Modeling
 - Database monitoring
 - DBMS
 - Developer Support

- Techniques
- Test in Lower Enviroments
 - Naming Standards
 - Scripts



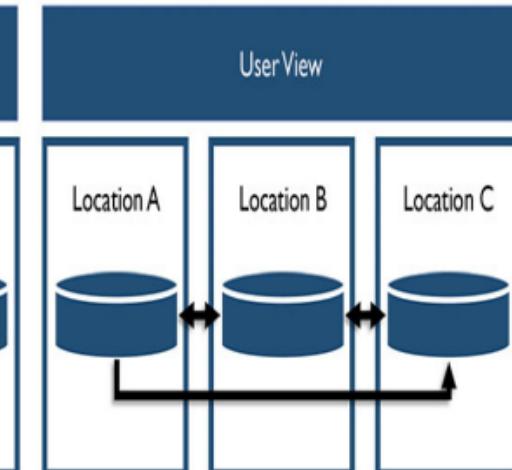
Activities

- Manage Database Technology
- Manage Databases
- Develop Database Instances
 - Physical
 - Access Controls
 - Storage
- PDM
- Load Data
- Replication
- Manage Database Performance
 - Availability
 - Execution
 - Service Levels
 - Enviroments
- Manage Test Data Sets
- Manage Data Migration

Centralized



Distributed, not Federated



CHAPTER 7: DATA SECURITY

Definition: Definition, planning, development and execution of security policies and procedures to provide proper authentication, authorisation, access and auditing of data and information assets



Goals

- Enable appropriate and prevent inappropriate access to enterprise data assets.
- Understand and comply with all the relevant regulations and policies for privacy, protection and confidentiality
- Ensure that the privacy and confidentiality needs of all stakeholders are enforced and audited

Business Drives

Risk Reduction - Business Growth - Security as an Asset

Principles

Collaboration - Enterprise Approach - Metadata Driven
Proactive Management - Clear Accountability
Reduce Risk by reducing exposure

Governance

Data Security Architecture
Implementation

- Readiness / Risk Assessment
- Organisation and cultural change
- Visibility into Entitlement
- Society and Outsourcing - Cloud Environment



Technology

Tools

- Antivirus / Encryption
- HTTPS / Data masking
- Identity Management
- IDS/IPS/Firewalls

Techniques

- CRUD Matrix
- Immediate Patch
- Metadata Repository
- Metrics / Projects
- Key MASKING



Essential Concepts

- Vulnerability
- Threat
- Risk Classification
 - Critical Risk Data
 - High Risk
 - Moderate Risk Data
- Data Security Organisation
 - NIST Risk Management Framework
- Security Processes
 - Access
 - Audit
 - Authentication
 - Authorisation
 - Entitlement
- Encryption
 - Hash
 - Private key
 - Public Key
- Obfuscation / Masking
 - Persistent Data Masking
 - In-flight
 - In-place
 - Dynamic
 - Methods
 - Substitution
 - Shuffling
 - Temporal Variance
 - Value Variance
 - NULLing/Deletion
 - Randomisation
 - Encryption
 - Expression Masking

Malware

- Adware
- Spyware
- Trojan Horse
- Virus
- Worm

Social / Phising

- ### Hacking
- White hat
 - Black hat

System Security Risks

- Excessive Privilege
- Legitimate privilege abuse
- Unauthorised privilege elevation
- Shared account
- Platform intrusion
- SQL injection
- Default Passwords
- Backup Data abuse

Types of Data Security Restriction

- Confidential
- Regulated
- Industry / Contract

Types of Data Security

- Facility
- Device
- Credential
- Electronic



Activities

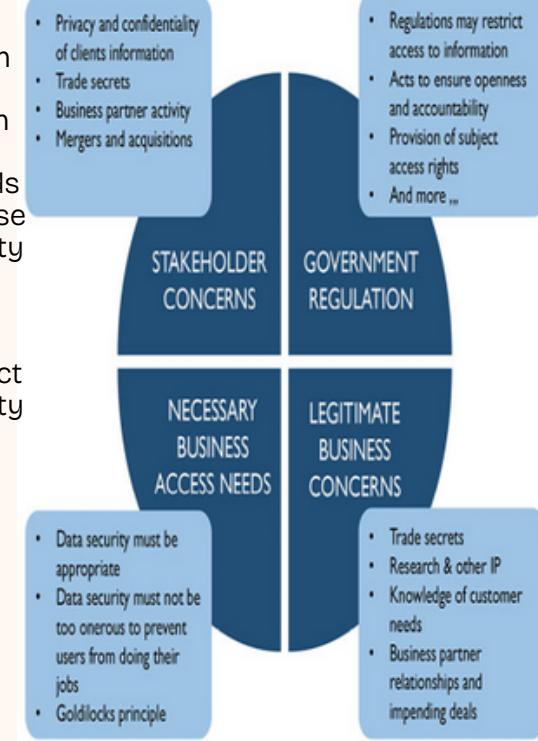
Identify Security Requirements

- Business
- Regulatory

Define Data Security Policy

Define Data Security Standards

- Confidentiality Levels
- Regulatory Categories
- Security Roles
- Assess Current Risks
- Implement Controls
- Manage and Maintain



CHAPTER 8: DATA INTEGRATION AND INTEROPERABILITY

Definition: Managing the movement and consolidation of data within and between applications and organisations



Goals

- Provide data securely, with regulatory compliance, in the format and timeframe needed.
- Lower cost and complexity of managing solutions by developing shared models and interfaces
- Identify meaningful events and automatically trigger alerts and actions
- Support business intelligence, analytics, master data management and operational efficiency efforts.



Essential Concepts

- ETL / ELT
Latency
 - Batch
 - CDC
 - Near-real-time
 - Asynchronous
 - Real-time Synchronous
 - Streaming

- Replication
Archiving
Canonical Model / EMF
Interaction Models
 - Point-to-point
 - Hub-and-spoke
 - Publish-subscribe
Data Exchange Standards

- DII Architecture
 - Application Coupling
 - Orchestration
 - Process Controls
 - EAI, ESB, SOA, CEP
 - Data Federation
 - Data Visualisation
 - Daas / Cloud-based



Business Drives

Manage Data Movement - Integrate Applications

Enterprise View of Df is Cost-Effective

Data Hubs - Standard Tools Reduce Complexity

Comply with Data Handling Regulations

KAs which DII is dependent on

Data Governance Data Security Data Storage
Data Architecture Metadata & Operations

DII is critical for DW & BI - Reference & Master Data



Activities

- Plan and Analyse
 - Define DII Lifecycle Requirements
 - Data Discovery
 - Data Linage
 - Profile Data
 - Business Rules

- Design DII Solutions
 - DII Architecture
 - Model
 - Mapping
 - Orchestration
Implement and Monitor

- Develop DII Solutions
 - Data Services
 - Data Flows
 - Data Migration Approach
 - Publication Approach
 - Complex Event Flows
 - Maintain DII Metadata



Governance

Data Sharing Agreements (MOU)

Data Linage
(Document - Metadata)

Metrics

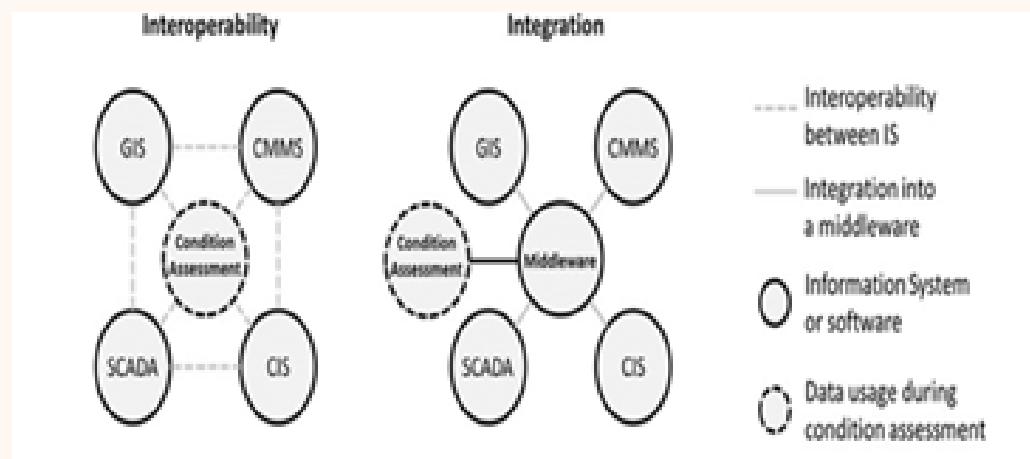
- Availability
- Volume
- Speed
- Latency
- Costs
- Complexity



Technology

Tools

- ETL
- Virtualisation Server
- ESB
- Business Rules Engine
- Modelling
- Profiling
- Metadata Repository



CHAPTER 9: DOCUMENT AND CONTENT MANAGEMENT

Definition: Planning implementation and control activities for lifecycle management of data and information in any form or medium



Goals

- To comply with legal obligations and customer expectations regarding records management
- To ensure effective and efficient storage, retrieval and use of documents and context.
- To ensure integration capabilities between structured and unstructured content



Essential Concepts

Content

- Document
- Management
- Metadata
- Modeling
- Delivery Methods

Controlled Vocabularies

- Management
- Term and Pick lists
- Synonym and Authority Lists
- Taxonomies
 - Flat
 - Hierarchical
 - Polyhierarchy
 - Facet
 - Network
 - Tagging
 - Thesauri
 - Ontology

Document Management

- Inventory
- Policy
- Classification
- Storage
- Retrieval
- Circulation
- Preservation
- Disposal

Digital Asset Management Data Map

- Inventory of ESI
- Application
- E-Discovery
 - Identification
 - Preservation
 - Collection
 - Processing
 - Review
 - Production
 - Presentation

Records Management

- Lifecycle
- Any Media
- VITAL Records
- Trustworthy Record
- Characteristics
 - Content
 - Context
 - Metadata
 - Timeliness
 - Permanency

Information Architecture

- Search Engine
- Semantic Model
- Semantic Search
- Unstructured Data
- Workflow

Government Laws and Regulations

What is the law?
Who enforces the legislation?
When? How? Where?

What regulatory rules and details are necessary to implement laws?

Policies and Standards

What is to happen? Where?
Why important?
Who is responsible?

Processes and Procedures

What needs to be done?
How to do it.

Work Instructions

What to do
Who will do the task?
Specific detailed steps for one task
Linked to a procedure

Other Documents that Express Evidence of Compliance (Records)

May include completed files, forms, tags, labels



Activities

Plan

- Lifecycle Management
- Records
- Content Strategy
- Policies
- Information Architecture

Manage

- Capture
- Version
- Backup
- Retention

Audit

Publish and Deliver



CHAPTER 10: REFERENCE AND MASTER DATA

Definition: Managing shared data to meet organisational goals, reduce risks associated with data redundancy, ensure higher quality, and reduce the cost of data integration



Goals

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



Essential Concepts

Reference Data

- Less Volatile
- Fewer Columns and Rows
- No Entity Resolution
- Domain Values and Definitions
- Storage
 - Code tables in DB
 - Linked with FKs
 - Reference Data Management System
- Structure
 - Lists
 - Cross-reference lists
 - Taxonomies
 - Ontologies
 - Internal Reference Data
 - Industry Reference Data
 - Geographic
 - Computational
- Standard Reference Data Set Metadata

Master Data

- System of record
- System of reference
- Trusted source
- Golden record
- Management
 - Requirements
 - Processing
 - Data Model Management
 - Data Acquisition
 - Data Validation
 - Standardisation
 - Enrichment
 - Entity Resolution
 - Matching
 - Identify Resolution
 - Data Sharing Stewardship
- Types
 - Party MD - Industry MD
 - Financial MD - Legal MD
 - Product MD - Location MD
- Data Sharing Architecture
 - Registry - Transaction Hub - Consolidated



Governance

Oversight and Metrics

Implementation

- Incremental - MD Architecture - Monitor
- RD Change Management
- Data Sharing Agreements
- Organisation Culture Change



Technology

Data Integration - Data Remediation - ODS - DSH
MDM Applications



Activities

MDM

- Define
 - Drivers
 - Requirements
 - Architecture
 - Stewardship
 - Maintenance
- Evaluate Data Sources
- Model Master Data
- Governance and Policies

RDM

- Define
 - Requirements
 - Drivers
 - Architecture
 - Stewardship
 - Maintenance
- Evaluate Data Sources
- Model Reference Data
- Governance and Policies



CHAPTER 11: BI AND DATA WAREHOUSING

Definition: Planning, implementation, and control process to provide decision support data and generate knowledge.

Goals

- To build and maintain the technical environment and technical and business intelligence activities
- To support and enable effective business analysis and decision making by knowledge workers

Technology

Tools

- Metadata Repository
 - Data dictionary
 - Business Glossary
 - Lineage
- Techniques
 - Prototype
 - Self-service BI - Audit data

Activities

Understand Requirements

- DW/BI Architecture
 - Warehouse
 - Technical
- DW/BI Management
- Implement BI
- Maintain Products

- Develop DW and Datamart
- Data
- Technology
- BI Tools
- Map Source to target
- Remediate
- Transform
- Populate the DW

Governance

Enable Business Acceptance
User Satisfaction
SLAs - Metrics and Reporting
Principles

- Business Goals
- End in Mind
- Think Globally, Build Locally
- Summarise Last
- Self-service - Collaborate
- Metadata - Built with the warehouse

Business Drivers

- BI support
- Compliance
- Operation support
- Decision support
- Innovation based on insights

Implementation

- Readiness / Risk Assessment
- Release Roadmap
- Configuration Management
- Organisation and culture change

Essential Concepts

BI: Type of Analysis

DW: Integrated Database, Software, EDW

Data Warehousing

- Immon
 - Relational Model
 - Subject Oriented
 - Integrated
 - Time Variant
 - Non-Volatile
 - Aggregate and Detail
 - Historical
- Kimball
 - Dimensional Model
 - Conformed Dimensions Bus

DW Architecture Concepts

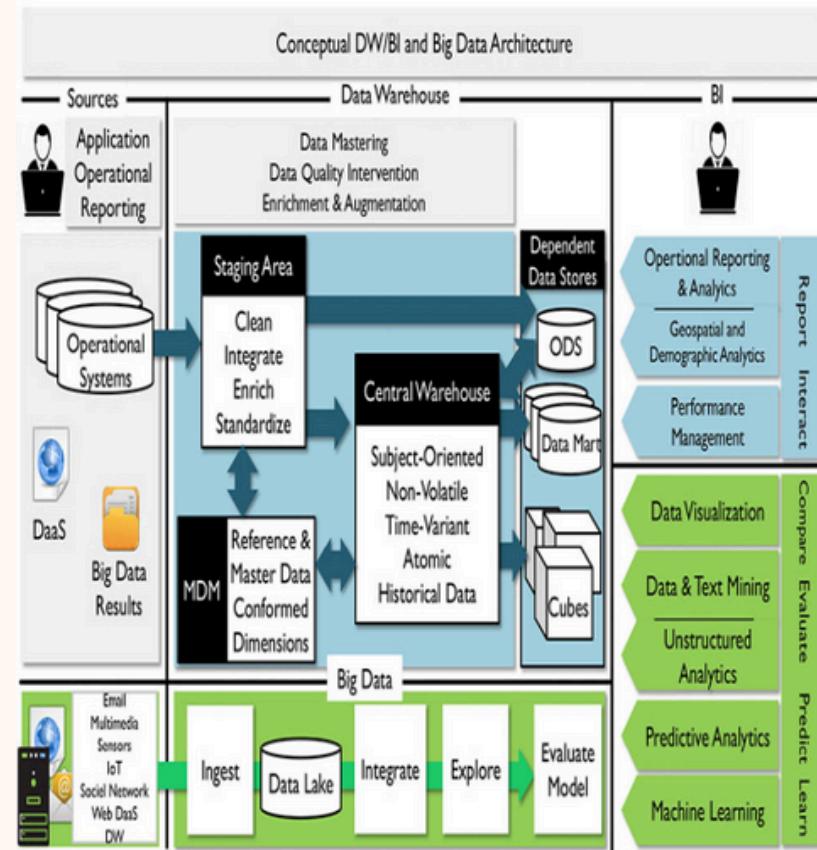
- Source Systems
- Data Integration
- Data Storage Areas (Staging, R and MD Conformed, Central Warehouse, ODS, Data Marts, Cubes (OLAP, ROLAP, MOLAP, HOLAP)

Load Processing

- Historical (Immon, Kimball, Data Vault)
- Batch CDC
- Real Time

Trickle Feeds, Messaging, Streaming

Conceptual DW/BI



CHAPTER 12: METADATA MANAGEMENT

Definition: Planning implementation and Control activities to enable access to high quality, integrated Metadata.

Goals

- Provide organisational understanding of business terms and usage
- Collect and integrate metadata from diverse sources
- Provide a standard way to access metadata
- Ensure Metadata quality and security

Governance

Process Controls - Documentation Standards - Guidelines - Metrics
Master Catalogue

Principles

Organisational - Commitment
Strategy - Protection - Access
Enterprise Perspective - Audit
Socialisation - Improvement
Implementation

- Risk/Readiness Assessment
- Organisational and Cultural

Technology

Tools Metadata repository
Techniques: Data lineage, Impact Analysis, Business and Technical Metadata for Big Data

Activities

- Define Metadata Strategy
- Define Metadata Architecture
- Create Metamodel
- Apply Metadata Standards
- Manage Metadata Standards
- Create and Maintain Metadata

Essential Concepts

Types

- Business, Technical, and operational MD Standard
- ISO 11179

Metadata for Unstructured Data

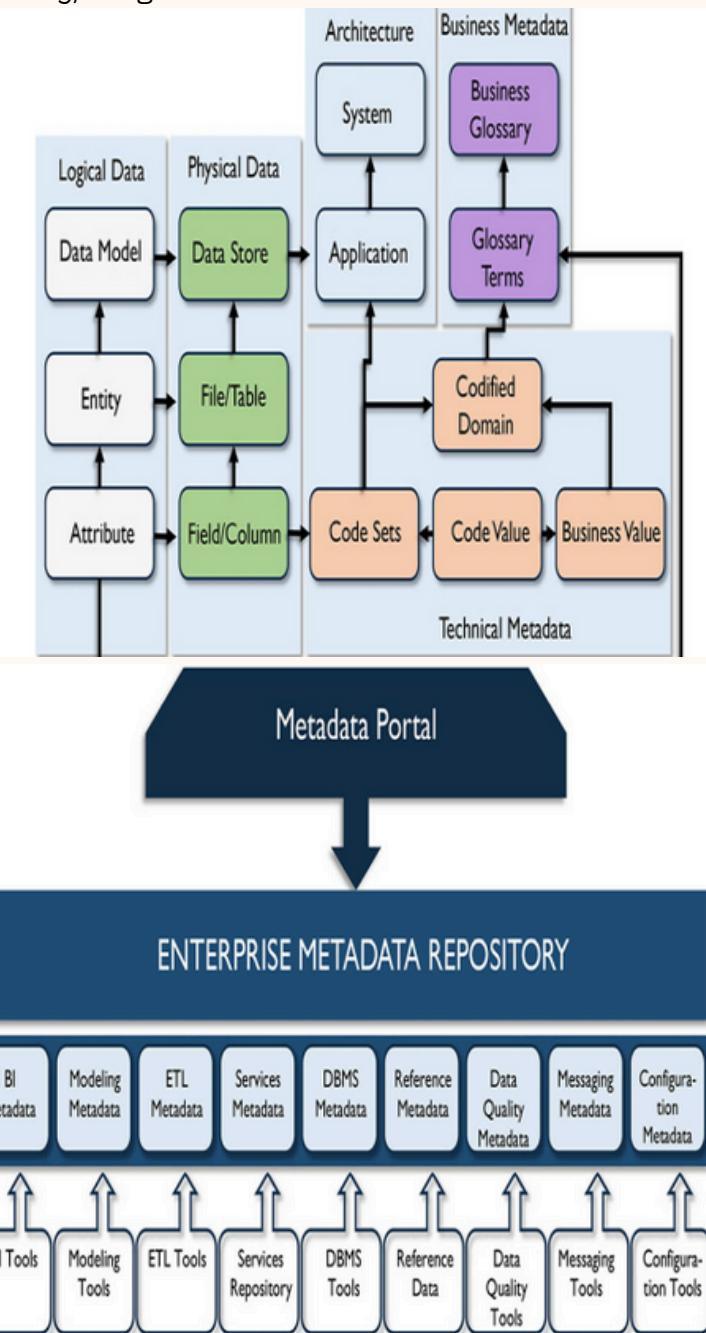
- Descriptive
- Structural
- Administrative
- Bibliographic
- Record Keeping Metadata
- Preservation Metadata

Sources of Metadata

- Applications
- Business Glossary
- CDM - Data dictionaries
- Data Integration Tools
- Datavase Management System Catalogues
- Data Mapping Tools
- Data Quality Tools
- Event Messaging Tools
- Modeling Tools
- Service Registries

Metadata Architecture

- Centralised (Copies - User Metadata)
- Distributed (Portal - No User Metadata)
- Hybrid
- Bi-directional



CHAPTER 13: DATA QUALITY

Definition: The planning, implementation and control of activities that apply quality management techniques to data, in order to assure it is fit for consumption and meets the needs of data consumers.



Goals

- Develop a governed approach to make data fit for purpose based on data consumers requirements.
- Define standards, requirements and specifications for data quality controls as part of the data lifecycle
- Define and implement processes to measure, monitor and report on data quality levels.
- Identify and advocate for opportunities to improve the quality of data through process and system improvements



Business Drives

Increase the value of Business Data

Reduce Poor Quality Data Risks

Organisational Efficiency and Productivity

Protect Organisations's Reputations

Principles

- Criticality - Lifecycle Management - Prevention
- Root Cause Remediation - Governance - SLA's
- Standards Driven - Transparency Measurement
- Embed in Business Processes
- Systematically Enforced

Governance

- Data Quality Policy - Metrics

Implementation Readiness and Cultural Change



Activities

Define High Quality Data and DQ Strategy

Identify Critical Data and Business Rules

Perform Initial Data Quality Assessment

Develop and Deploy DQ Operations: Manage DQ Rules, Monitor, Data elements, record, set, issues, DQ SLA's Reporting



Technology

Tools

- Data Profiling
- Data Querying
- Modelling
- ETL / Rules Templates
- Metadata Repositories

Techniques

- Preventative Actions
 - Data Entry Controls - Training
 - Define and Enforce Rules
 - Demand High Quality Data
 - Implement DQ and Stewardship
 - Formal Change Control
- Corrective Actions
 - Automated Correction
 - Manually-driven Collection
 - Manual Correction
- Quality Check and Audit Code Modules
- Effective DQ Metrics
- Statistical Process Control
 - Common Causes (Inherent in process and In Statistical Control)
 - Special Causes: Unpredictable
- Root Cause Analysis
 - Pareto 80/20 Rule
 - Fishbone / Ishikawa cause
 - Track and trace
 - Process Analysis
 - Five Whys
- Data Profiling
- DQ and Data Processing



Essential Concepts

Data Quality

- Meets Expectations
- Fit for Purpose

Critical Data

- Required by
 - Regulatory Reporting
 - Financial Reporting
 - Business Policy
 - Ongoing Operations
 - Business Strategy
 - Master Data

DQ Dimensions

- Accuracy
- Completeness
- Integrity
- Reasonability

- Timeliness
- Uniqueness
- Validity

DQ Metadata

- Defines Data
- Results of DQ Measurements
- Repository

DQ Standard ISO 8000

DQ Improvement Lifecycle

- Shewhart/Deming cycle
 - Plan, Do, Check, Act

DQ Business Rule Types

- Describe how data must exist to be useful
- Implemented in
 - Software
 - Data Entry Templates

Causes of DQ Issues

- Lack of Leadership
 - Data Entry, Interface Issues, List Entry Placement, Field Overloading, Training Issues, Business Process Changes, Inconsistent Process Execution



CHAPTER 14: BIG DATA AND DATA SCIENCE

Definition: The collection (Big Data) and analysis (Data Science, Analytics and Visualisation) of many different types of data to find answers and insights for questions that are not known at the start of analysis.



Goals

- Discover relationships between data and business
- Support the iterative integration of data sources into the enterprise
- Discover and analyse new factors that might affect the business
- Publish data using visualisation techniques in an appropriate, trusted and ethical manner



Governance

- Visualisation Channel Management
Data Science / Visualisation Standards
Data Security, Metadata, Data Quality and Metrics Implementation
- Strategy Alignment
 - Readiness / Risk Assessment
 - Business Relevance - Business Readiness
 - Economic Viability - Prototyping
- Principles
- Manage Big Data Metadata
 - Inventory, Origins and Value



Technology

- Tools
- MPP, Distributed File Databases, In-database Algorithms
 - Big Data Cloud Solutions, Statistical and Graphical
 - Data Visualisation, Tools R, Python
- Techniques
- Analytic (Types of Analytics)
 - Big Data Modeling



Activities

- | | |
|--------------------------|--------------------------------|
| Define Big Data Strategy | Integrate / Align for Analysis |
| Choose Data Sources | Explore Using Models |
| Acquire and Ingest | Deploy and Monitoring |
| Develop Hypotheses | |



Essential Concepts

- Data Science
- Rich Data Sources
 - Information Alignment and Analysis
 - Information
 - Presentation of
- Data science process
- Big Data
- Volume, Velocity, Variety
 - Viscosity, Volatility, Veracity
- Big Data Architecture Components
- Data Lake
 - Mine Data, Raw, Alternative Data warehouse storage, Online records archive, streaming Data, Pattern Recognition
 - Mange Metadata
- Types of Analysis
- Machine Learning
 - Supervised, Unsupervised, Reinforcement
 - Sentiment Analysis
 - Keywords, Natural Language Processing (NLP)
 - Mining
 - Data Reveal patterns, Unknown Outcomes / Text Automatically Classify
 - Predictive and Prescriptive
 - Supervised Probability Models / Implications of Decisions
 - Unstructured: Scanning and Tagging
 - Operational: BI and Streaming / Data visualisation / Mashups
- BI vs Data Science
- Business Intelligence
 - Descriptive. Hindsight, Based on, What
 - Data Science
 - Predictive. Insight. Based on predictive models, what will happen
 - Prescriptive. Foresight, based on what can we do to make things
- Services-Based Architecture
- Lambda Architecture
 - Batch Layer
 - Data Lake
 - Historical Data
 - Structure over time
 - Inserts
 - Speed Layer
 - Real Time Data, ODS
 - Inserts and Updates
 - Serving Layer

