

COURSEWARE

DATA MANAGEMENT COURSEWARE BASED ON CDMP FUNDAMENTALS

SECOND REVISED EDITION



Data Management courseware based on
CDMP Fundamentals

Second revised edition

Colophon

Title: Data Management courseware based on CDMP Fundamentals –
Second revised edition

Developed with Bas van Gils, Ingrid Stap and Denise Harders

Authors: Strategy Alliance BV / And More Group BV A.O.

Illustration cover: &More group, Denise Harders

Publisher: Van Haren Publishing, 's-Hertogenbosch

ISBN Hard Copy: 978 94 018 1 291 7

Edition: First edition, first print, August, 2021
Second edition, first print, February, 2024
Third edition, first print, March, 2025

Design: Van Haren Publishing, 's-Hertogenbosch

Copyright: © Van Haren Publishing 2025

For further information about Van Haren Publishing please e-mail
us at: info@vanharen.net or visit our website: www.vanharen.net

This material contains diagrams and text Information based upon: DAMA NL materials.
All rights reserved.

No part of this publication may be reproduced in any form by print, photo print, microfilm
or any other means without written permission by the publisher.

Although this publication has been composed with much care, neither author, nor editor,
nor publisher can accept any liability for damage caused by possible errors and/or
incompleteness in this publication.

Material in this document has been sourced from Data Management: a gentle
introduction. No part of this document may be reproduced in any form without the
written permission of both Van Haren Publishing.

Publisher about the Courseware

The Courseware was created by experts from the industry who served as the author(s) for this publication. The input for the material is based on existing publications and the experience and expertise of the author(s). The material has been revised by trainers who also have experience working with the material. Close attention was also paid to the key learning points to ensure what needs to be mastered.

The objective of the courseware is to provide maximum support to the trainer and to the student, during his or her training. The material has a modular structure and according to the author(s) has the highest success rate should the student opt for examination. The Courseware is also accredited for this reason, wherever applicable.

In order to satisfy the requirements for accreditation the material must meet certain quality standards. The structure, the use of certain terms, diagrams and references are all part of this accreditation. Additionally, the material must be made available to each student in order to obtain full accreditation. To optimally support the trainer and the participant of the training assignments, practice exams and results are provided with the material.

Direct reference to advised literature is also regularly covered in the sheets so that students can find additional information concerning a particular topic. The decision to leave out notes pages from the Courseware was to encourage students to take notes throughout the material.

Although the courseware is complete, the possibility that the trainer deviates from the structure of the sheets or chooses to not refer to all the sheets or commands does exist. The student always has the possibility to cover these topics and go through them on their own time. It is recommended to follow the structure of the courseware and publications for maximum exam preparation.

The courseware and the recommended literature are the perfect combination to learn and understand the theory.

-- Van Haren Publishing

Other publications by Van Haren Publishing

Van Haren Publishing (VHP) specializes in titles on Best Practices, methods and standards within four domains:

- IT and IT Management
- Architecture (Enterprise and IT)
- Business Management and
- Project Management

Van Haren Publishing is also publishing on behalf of leading organizations and companies: ASLBiSL Foundation, BRMI, CA, Centre Henri Tudor, Gaming Works, IACCM, IAOP, IFDC, Innovation Value Institute, IPMA-NL, ITSqc, NAF, KNVI, PMI-NL, PON, The Open Group, The SOX Institute.

Topics are (per domain):

IT and IT Management

ABC of ICT
ASL®
CATS CM®
CMMI®
COBIT®
e-CF
ISO/IEC 20000
ISO/IEC 27001/27002
ISPL
IT4IT®
IT-CMF™
IT Service CMM
ITIL®
MOF
MSF
SABSA
SAF
SIAM™
TRIM
VeriSM™

Enterprise Architecture

ArchiMate®
GEA®
Novius Architectuur
Methode
TOGAF®

Business Management

BABOK® Guide
BiSL® and BiSL® Next
BRMBOK™
BTF
EFQM
eSCM
IACCM
ISA-95
ISO 9000/9001
OPBOK
SixSigma
SOX
SqEME®

Project Management

A4-Projectmanagement
DSDM/Atern
ICB / NCB
ISO 21500
MINCE®
M_o_R®
MSP®
P3O®
PMBOK® Guide
Praxis®
PRINCE2®

For the latest information on VHP publications, visit our website: www.vanharen.net.

Intro to Data Management courseware based on CDMP Fundamentals

More and more organisations see 'data' as the fuel on which the business engine runs. Themes such as data-driven work and smart solutions with big data and artificial intelligence are relevant in all sorts of sectors. This development means that more attention is being paid to data management: what does it mean to manage data as an 'asset'? And how do we guard the balance between 'grip on data' on the one hand, and 'value creation with data' on the other?

DAMA is the international professional organisation in the field of data management. The Data Management Body of Knowledge (DMBOK) is the best known publication, and Certified Data Management Professional (CDMP) the best known certification. The purpose of this training course is to prepare for the CDMP exam. The training covers all relevant parts of the DMBOK and contains besides theory also a number of practical exercises and practice questions which prepare for the exam.

Literature reference

The chapter structure of this courseware and the recommended Data Management Body of Knowledge (DMBOK) has been made alike. Therefore if you are looking for additional references you can do so in the DMBOK.

Table of content

| | <i>--- Slide number</i> | <i>--- Page slidedeck</i> | <i>--- Page notesdeck</i> |
|--|-------------------------|---------------------------|---------------------------|
| Reflection | | 8 | |
| Agenda | | 10 | |
| Data Management | 1 | 11 | 83 |
| Data(Management) strategy | 11 | 16 | 95 |
| Data Governance | 15 | 18 | 99 |
| Governance vs. Management | 17 | 19 | 101 |
| KPI dashboard | 22 | 21 | 107 |
| Data Architecture | 27 | 24 | 113 |
| Architecture | 29 | 25 | 115 |
| Architecture scope | 30 | 25 | 116 |
| Frameworks | 31 | 26 | 117 |
| Enterprise data model | 32 | 26 | 118 |
| Data in motion & data at rest | 33 | 27 | 119 |
| Exercise: Data Governance & Data Architecture | 34 | 27 | 120 |
| Data Modeling & Design | 40 | 30 | 126 |
| Business drivers | 42 | 31 | 128 |
| Abstraction levels of data modeling | 43 | 32 | 129 |
| Languages & schemas | 45 | 33 | 132 |
| Tools & techniques | 47 | 33 | 134 |
| Data Storage & Operations | 53 | 36 | 140 |
| DBA monitoring | 55 | 37 | 142 |
| Data Security | 65 | 42 | 152 |
| Data Integration & Interoperability | 78 | 49 | 168 |
| Virtualization | 82 | 51 | 172 |
| Document & Content Management | 90 | 55 | 180 |
| Reference & Master Data Management | 98 | 59 | 189 |
| Reference Data | 100 | 60 | 191 |
| Master Data | 101 | 60 | 192 |
| In Practice | 103 | 61 | 194 |
| Data Warehousing & BI | 109 | 64 | 200 |
| Business intelligence | 111 | 65 | 202 |
| Data Warehouse Architectures | 112 | 66 | 204 |
| Metadata Management | 118 | 69 | 211 |

| | | | |
|--------------------------------|-----|----|-----|
| Types of metadata | 120 | 70 | 213 |
| Relationship with models | 121 | 70 | 214 |
| Sources of metadata | 122 | 71 | 215 |
| Metadata architecture | 123 | 71 | 216 |
| Data Quality Management | 131 | 75 | 225 |
| Quality Dimensions | 134 | 76 | 229 |
| Process in practice | 137 | 76 | 232 |
| DQ Monitor | 138 | 77 | 234 |
| Exam | 143 | 79 | 239 |
| Practice Exams Info | 246 | | |

Self-Reflection of understanding Diagram

‘What you do not measure, you cannot control.’ – Tom Peters

Fill in this diagram to self-evaluate your understanding of the material. This is an evaluation of how well you know the material and how well you understand it. In order to pass the exam successfully you should be aiming to reach the higher end of Level 3. If you really want to become a pro, then you should be aiming for Level 4. Your overall level of understanding will naturally follow the learning curve. So, it’s important to keep track of where you are at each point of the training and address any areas of difficulty.

Based on where you are within the Self-Reflection of Understanding diagram you can evaluate the progress of your own training.

| <i>Level of Understanding</i> | <i>Before Training (Pre-knowledge)</i> | <i>Training Part 1 (1st Half)</i> | <i>Training Part 2 (2nd Half)</i> | <i>After studying / reading the book</i> | <i>After exercises and the Practice exam</i> |
|--|--|-----------------------------------|-----------------------------------|--|--|
| <i>Level 4 I can explain the content and apply it .</i> | | | | | |
| <i>Level 3 I get it! I am right where I am supposed to be.</i> | | | | | <i>Ready for the exam!</i> |
| <i>Level 2 I almost have it but could use more practice.</i> | | | | | |
| <i>Level 1 I am learning but don't quite get it yet.</i> | | | | | |

(Self-Reflection of Understanding Diagram)

Write down the problem areas that you are still having difficulty with so that you can consolidate them yourself, or with your trainer. After you have had a look at these, then you should evaluate to see if you now have a better understanding of where you actually are on the learning curve.

Troubleshooting

Problem areas:

Topic:

Part 1

Part 2

You have gone
through the book
and studied.

You have answered
the questions and
done the practice
exam.

Timetable

Day 1

| Time: | Subject: |
|--------------|---|
| +/- 15 min.: | Walk-in |
| +/- 60 min.: | Intro Data Management + exercise Maturity |
| +/- 20 min.: | Data Governance. |
| +/- 15 min.: | break of 10 minutes (in reality 15) |
| +/- 30 min.: | Data Architecture |
| +/- 30 min.: | Exercise DG+DA |
| +/- 60 min.: | Lunch |
| +/- 60 min.: | Data modeling & Design |
| +/- 20 min.: | Data Storage & Operations |
| +/- 15 min.: | break of 10 minutes (in reality 15) |
| +/- 20 min.: | Data Security |
| +/- 20 min.: | Data Integration & Interoperability |

Day 2

| Time: | Subject: |
|--------------|-------------------------------------|
| +/- 15 min.: | Walk-in |
| +/- 20 min.: | Document & Content |
| +/- 20 min.: | Reference & Master data |
| +/- 60 min.: | Data Warehouse & BI |
| +/- 15 min.: | break of 10 minutes (in reality 15) |
| +/- 15 min.: | Exercise Data warehouse & BI |
| +/- 20 min.: | Metadata |
| +/- 60 min.: | Lunch |
| +/- 60 min.: | Data Quality |
| +/- 15 min.: | Exercise data quality |
| +/- 15 | break of 10 minutes (in reality 15) |
| +/- 60 min.: | Exam training |

DATA MANAGEMENT

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.

© Van Haren Publishing 1

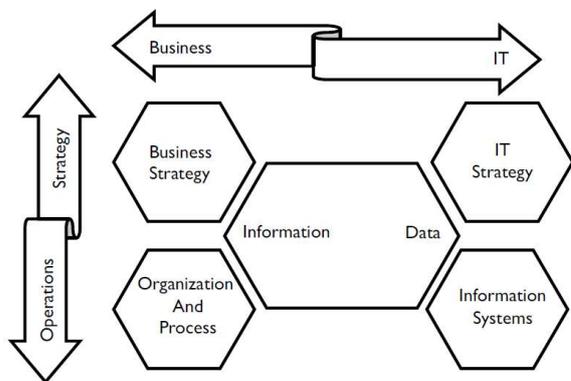
Framework

DAMA-DMBOK Wheel
Data Management Body Of Knowledge (DMBOK) of the Data Management Association International (DAMA)

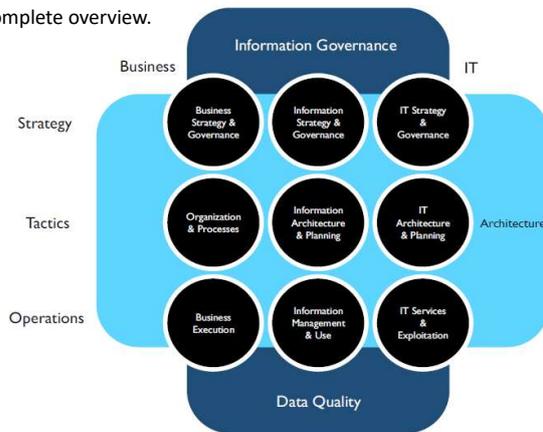
© Van Haren Publishing 2

Other data management frameworks

See the section on data management frameworks in Chapter 1 for a complete overview.



Strategic Alignment Model – SAM
Figures taken from: DAMA DMBOK v2



Amsterdam information model – AIM
Figures taken from: DAMA DMBOK v2

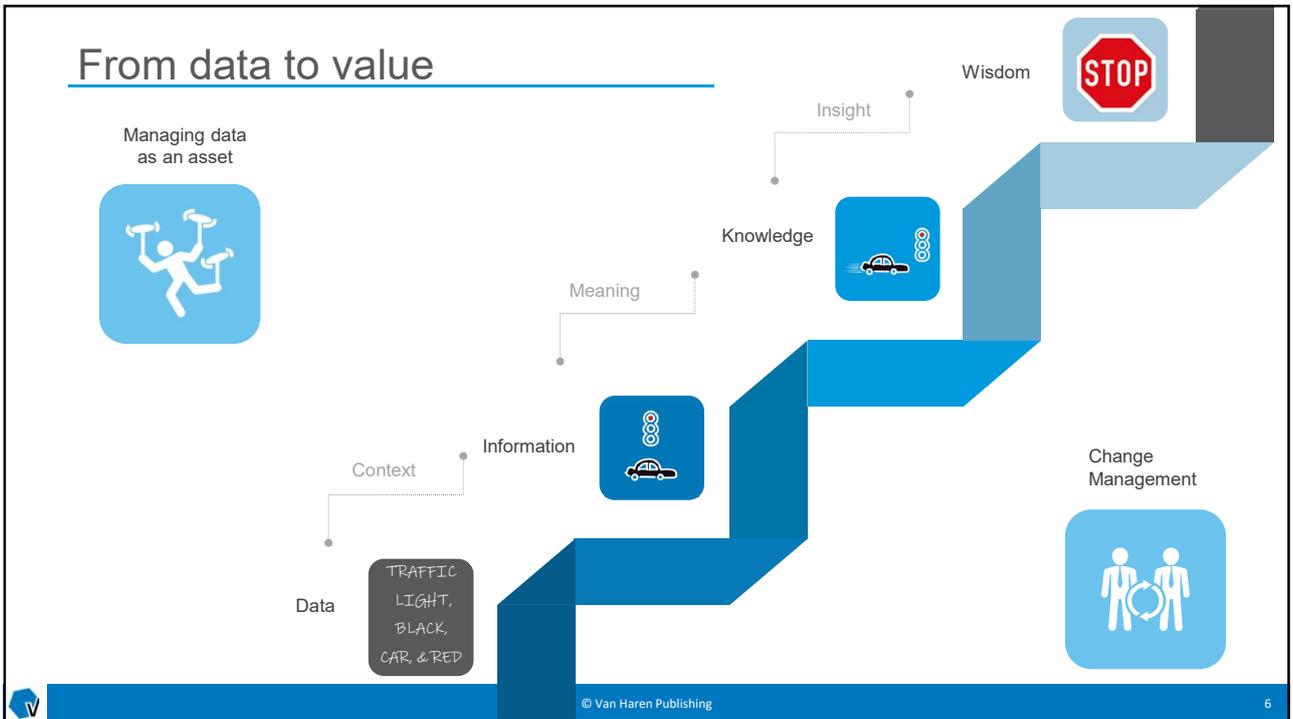
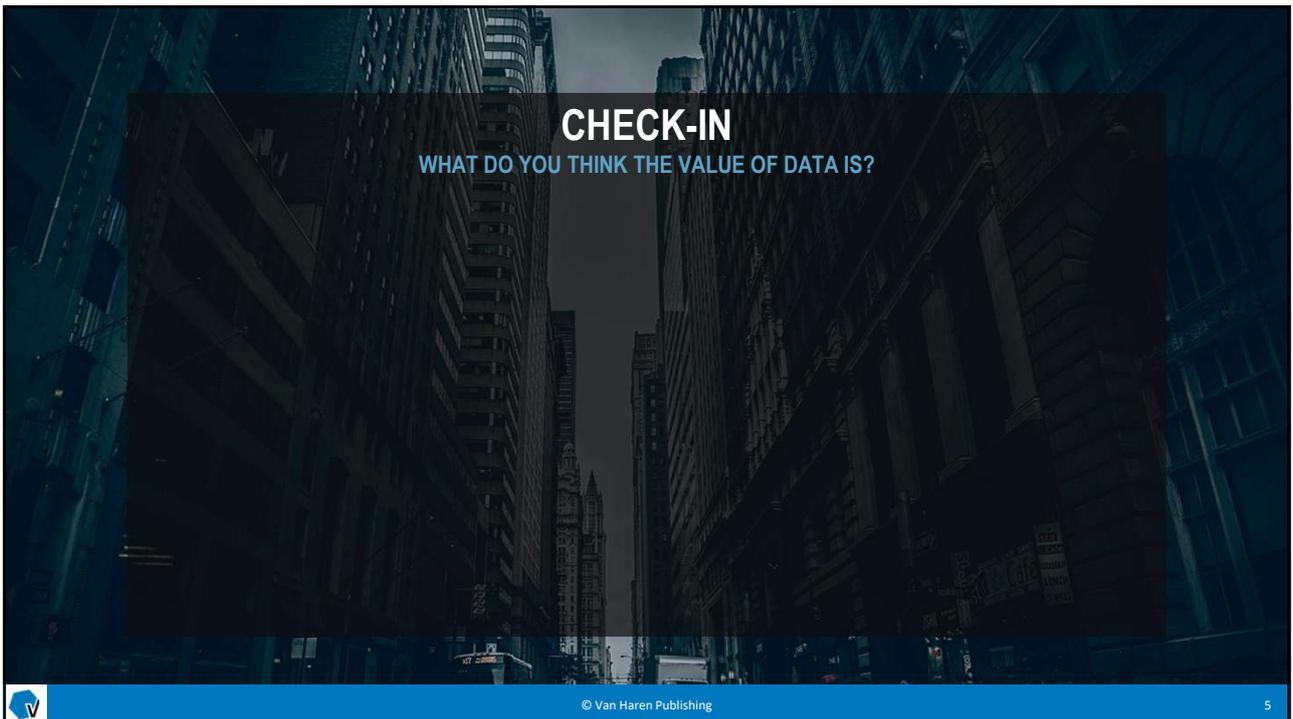


What are we going to do per knowledge area?



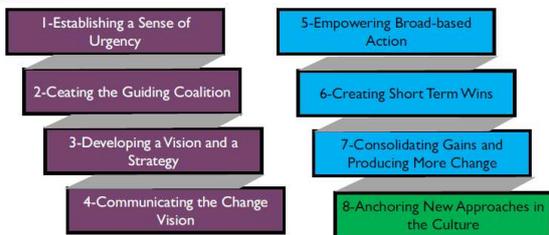
- 1 Intro
- 2 Theory – DAMA Framework
- 3 In practice
- 4 Exercises/homework





Examples of change management frameworks

8-steps of Kotter



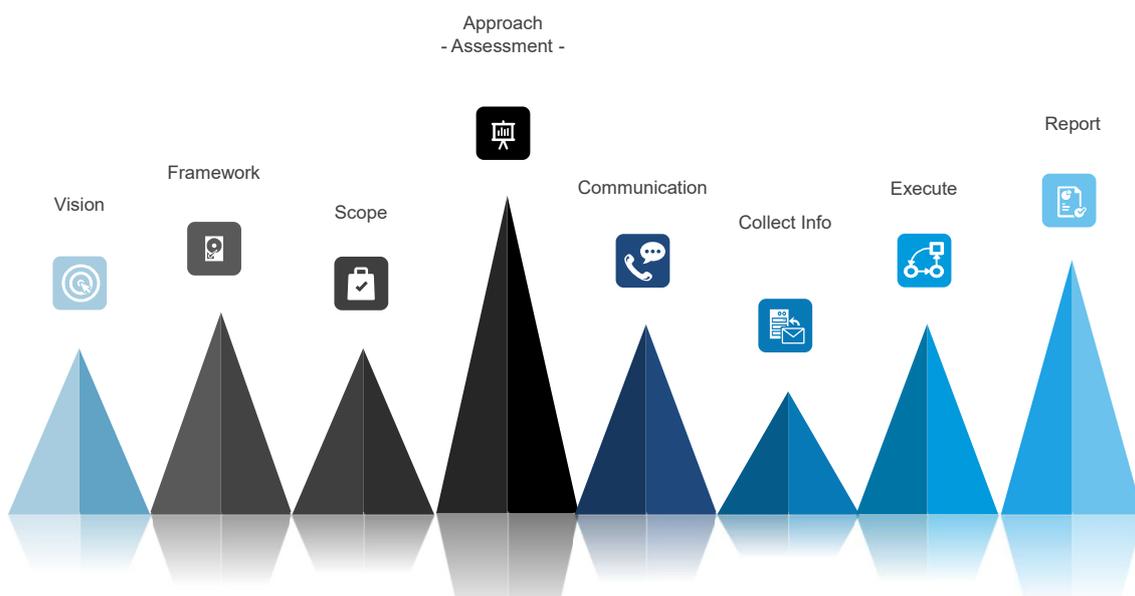
ADKAR



Figure taken from: DAMA DMBOK v2



Where do you start?



Maturity levels: exercise

| 0: Absence of capability | 1: Initial/Ad-hoc | 2: Repeatable | 3: Defined | 4: Managed | 5: Optimized/Data-driven |
|--------------------------|-------------------|---------------|------------|------------|--------------------------|
| | | | | | |



DAMA wheel evolved

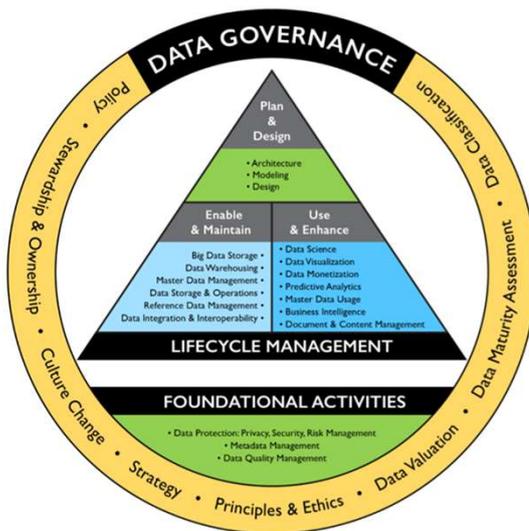


Figure taken from: DAMA DMBOK v2



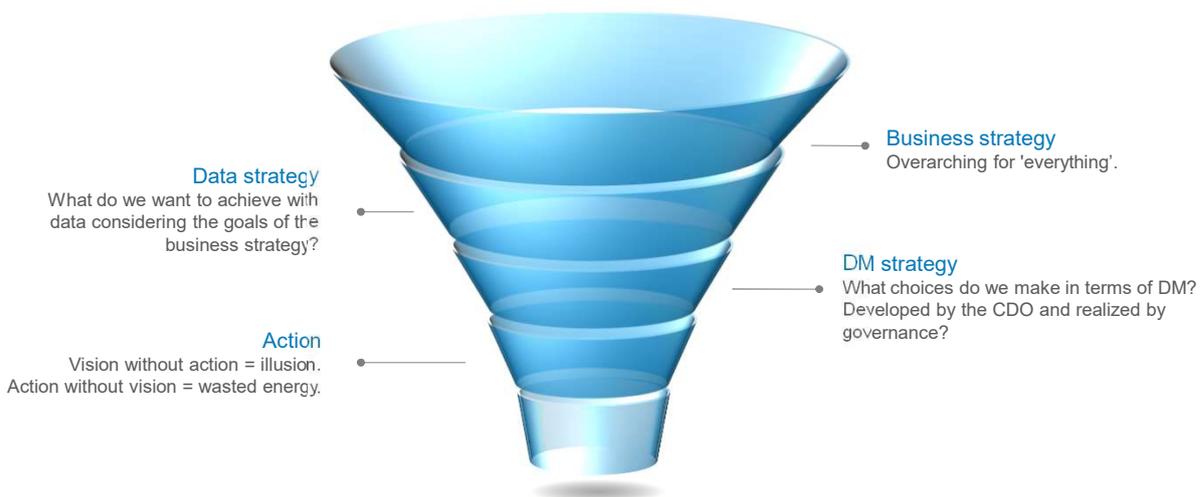
Data (management) strategy

Strategy

Set of choices as input for (making) a strategic plan.

Strategic plan

High-level planning for achieving strategic objectives.



Practice questions

1. Which one of the following is NOT true when describing Capability Maturity Model Integration (CMMI)?

- A. Model framework to assess data and process maturity.
- B. Model framework to determine priorities.
- C. Model framework to institute process and data improvement.
- D. Defines six levels of process maturity.

2. Data management is:

- A. An ongoing initiative.
- B. A one-off activity.
- C. Something that you can do alone.
- D. Easy to implement and will take less than a week.
- E. All but A is correct.



KNOWLEDGE AREA

DAMA area and definition

Context diagram

Which part of the DAMA circle?

How much % it counts for the exam

DEFINITION OF KNOWLEDGE AREA

GENERIC CONTEXT DIAGRAM

Definition: High-level description of the knowledge area

Goals: Purposes of the Knowledge Area

- Goal 1
- Goal 2

Business Drivers

Inputs:

- Input 1
- Input 2
- Input 3

Inputs are generally outputs from other Knowledge Areas

Activities:

1. Planning Activity / Activity Group (P)
 1. Sub activity
 2. Sub activity
2. Control Activity / Activity Group (C)
3. Development Activity / Activity Group (D)
4. Operational Activity / Activity Group (O)

Suppliers:

- Supplier 1
- Supplier 2

Participants:

- Role 1
- Role 2

Consumers:

- Role 1
- Role 2

Technical Drivers

Techniques:

- Methods and procedures to execute activities

Tools:

- Software package types to support activities

Metrics:

- Measurable results of the process

(P) Planning, (C) Control, (D) Development, (O) Operations

Figure taken from: DAMA DMBOK v2

© Van Haren Publishing 13

CHECK-IN

DATA GOVERNANCE & DATA MANAGEMENT ARE THE SAME

Agree ————— Disagree

© Van Haren Publishing 14

DATA GOVERNANCE

DATA GOVERNANCE IS EXERCISE OF AUTHORITY AND CONTROL OVER THE MANAGEMENT OF DATA ASSETS.

DATA GOVERNANCE IS A DISCIPLINE OF CATALOGING AND DEFINING IMPORTANT DATA, ASSIGNING OWNERSHIP OF DATA AND INCORPORATING THE MANAGEMENT OF DATA INTO THE DAILY BUSINESS PROCESSES.

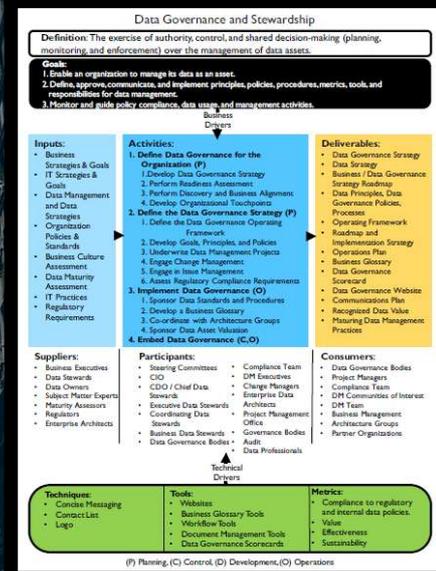


Figure taken from: DAMA DMBOK v2



Topics



Data governance vs. data management

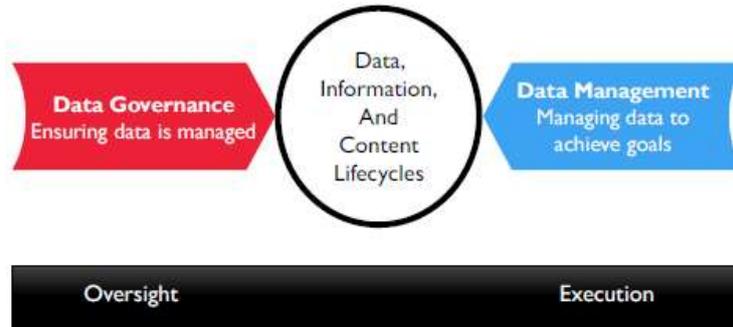


Figure taken from: DAMA DMBOK v2



DG organization parts

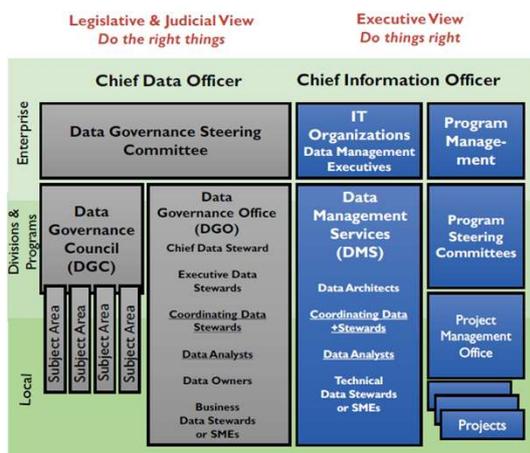
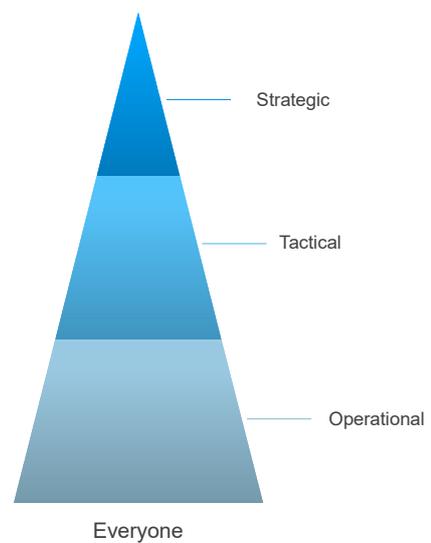
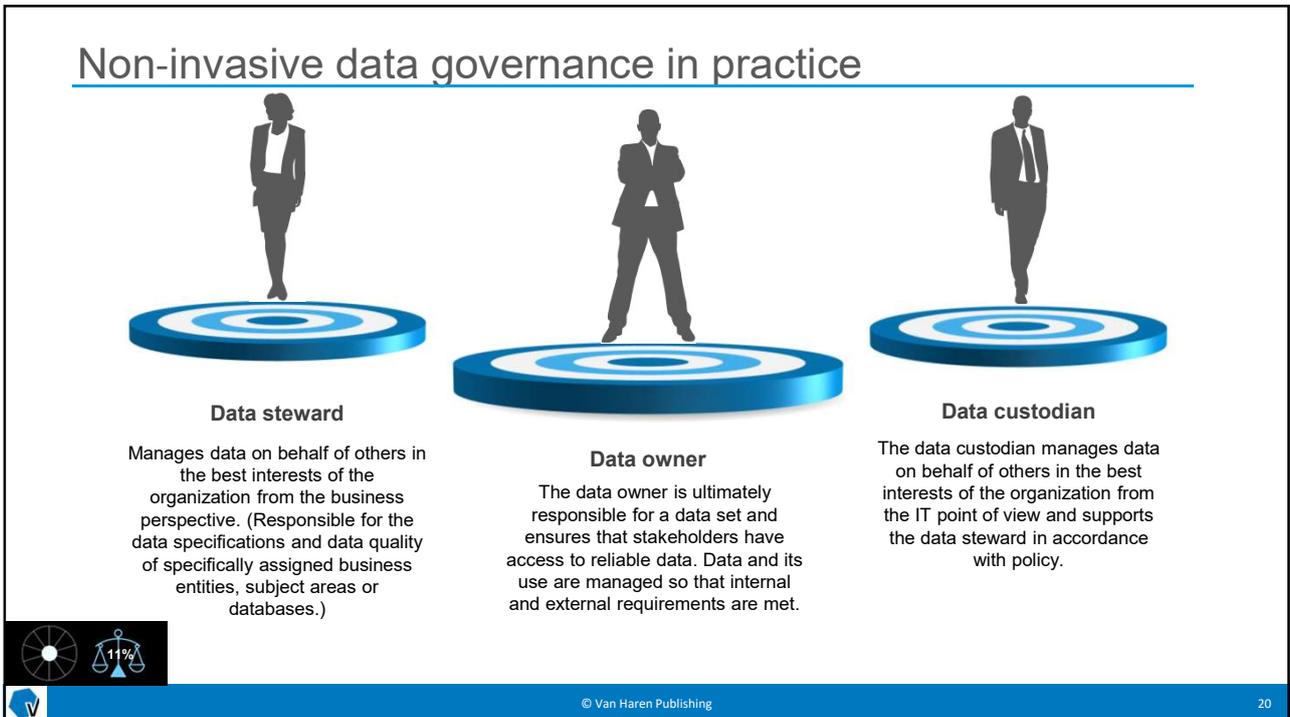
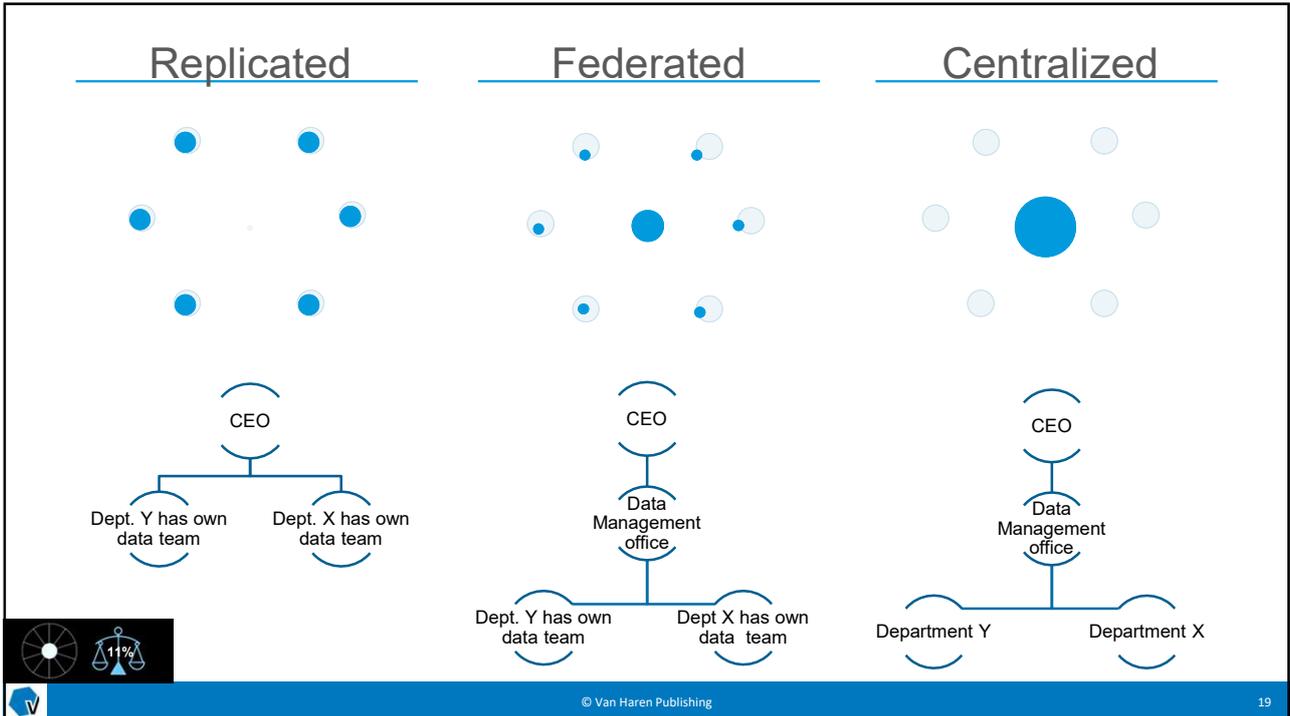


Figure taken from: DAMA DMBOK v2

Who is responsible?



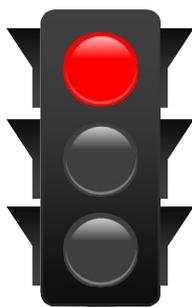


Increase data quality & integrate DMI process in practice

| Key | Summary | Issue Type | Status | Assignee | Due Date | Linked Issues | Description | Security Level | DGB Prio | Data Domain | DAMA Category | Owner DGB | Impact score |
|--------|--|-----------------------|-------------|--------------|----------|----------------|---|----------------|----------|-------------|-------------------------|-------------|--------------|
| EX-728 | Consumer is not offered a new rate in accordance with terms and conditions | Data Management Issue | Review | XXXX, Willem | | | It seems that a (large) number of subscription due dates are not on par with current terms. | Internal | Normal | Backoffice | Data Quality | XXXX, Frank | 12 |
| EX-720 | Consumers and agents are stored in multiple places (systems) | Data Management Issue | In Progress | XXXX, Remko | | EX-464, EX-142 | MTSD nr.: CR864392 Jira RFC: EX-464 Status = PO Analysis but unassigned | Internal | Normal | Backoffice | Reference & Master Data | XXXX, Frank | 14 |
| EX-954 | Insurance companies regularly merge, addresses no longer correct; | Data Management Issue | NEW | Unassigned | | | | Internal | Normal | Backoffice | Data Quality | XXXX, Frank | |



KPI dashboard



Define and establish quality criteria for new or existing data elements.

= <90%



Analyze reports from the data quality tool.

= 90% - 95%



High customer satisfaction by proactively reporting and addressing data issues.

= >80%



Case: Data management at a supermarket chain - DG



Context:

- Large supermarket listed on AEX.
- 895 stores in the Netherlands.
- 800 owned stores and 95 franchises.
- We know there is a lot of data, we just don't know where it is.



Setting:

- You are responsible for the purchasing, marketing and sales of apples.
- You discover that you have trouble finding the right data to manage your business.
- You want to "do more with data".
- You want to improve logistics, better service to customers (still unclear what that means).
- No data governance function exists as yet.
- A small group of your colleagues is motivated to get started.



Challenge:

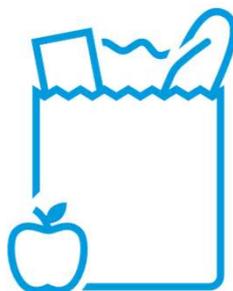
- What data governance model would you choose for the supermarket organization (centralized/decentralized, or hybrid) and why?
- When you have resolved this and have time left:
 - Describe your roles and responsibilities from the position of data management.
 - Are you a data steward, data custodian and/or the data owner?



Case: Data management at a supermarket chain - DG

Organization

- What type of management is in place, replicated/federated/centralized?
- Which employees deal with data?
- Which data roles can be defined?



Who is responsible ?

- Is ownership assigned?
- In what way is data ownership organized? (Department level, system level, product level.)
- Do the various roles have decision-making authority?



Practice questions

1. Which one of the following is NOT a part of a Data Management Plan?

- A. Describe the roles and resources of program staff.
- B. Define future direction of data management activities in a work plan.
- C. Implement facilities and tools for managing metadata resources.
- D. Development of a quality management plan.

2. Which of these best describes the relationship between Data Governance and Data Management?

- A. Data Governance is ensuring data is managed, whereas Data Management involves managing data to achieve business goals.
- B. Data Management is ensuring data is managed, whereas Data Governance involves managing data to achieve business goals.
- C. Data Governance is an IT-led initiative, whereas Data Management is a business function.
- D. Data Governance and Data Management both mean the same thing.
- E. Data Governance is separate from Data Management.



CHECK-IN

DATA ARCHITECTURE IS MAINLY CONCERNED WITH...?

| | | |
|----------|------------------|--------------|
| Design | | |
| Describe | | |
| | Boxes and arrows | Other things |



DATA ARCHITECTURE

DATA ARCHITECTURE REFERS TO THE SCIENCE OF BUILDING STRUCTURES AND THEIR RESULTS. IN A MORE GENERAL SENSE, ARCHITECTURE REFERS TO AN ORGANIZED ARRANGEMENT OF COMPONENTS AND ELEMENTS INTENDED TO OPTIMIZE THE FUNCTION, PERFORMANCE FEASIBILITY, COST, AND AESTHETICS OF THE ENTIRE STRUCTURE/SYSTEM

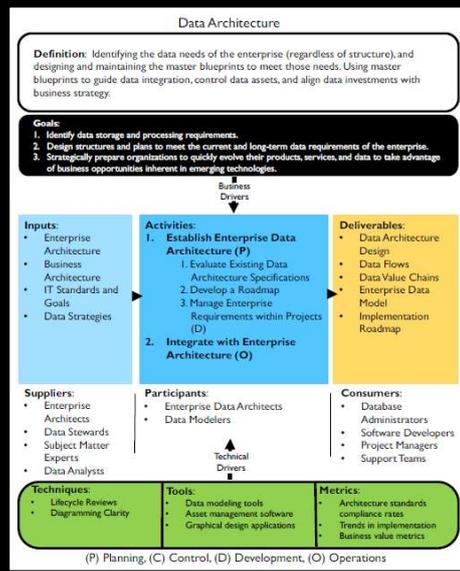
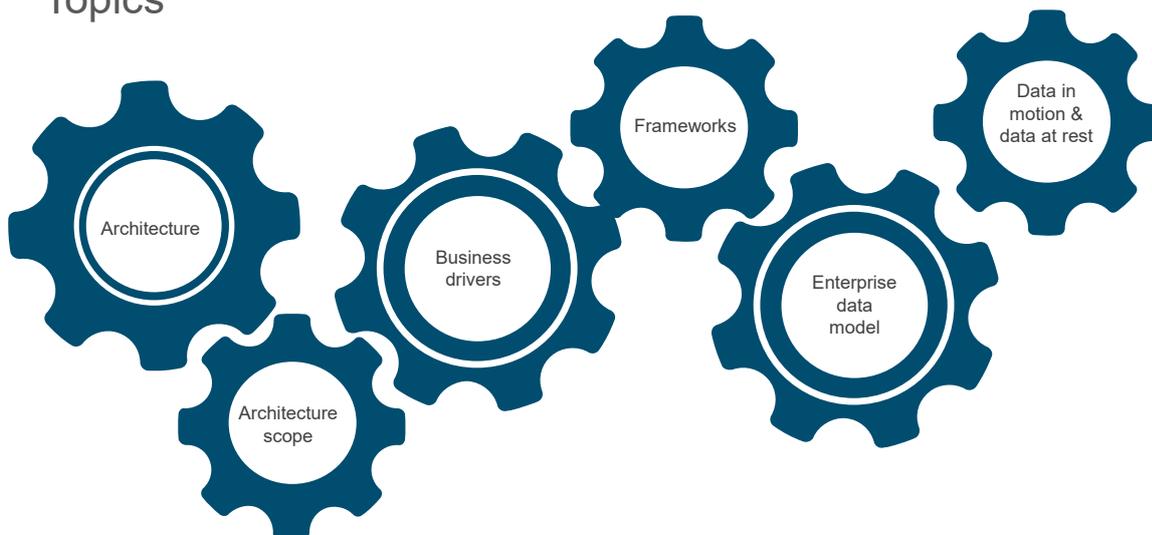


Figure taken from: DAMA DMBOK v2



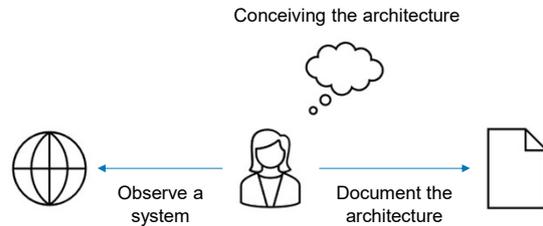
Topics



Architecture

Architecture of a system:

- (1) Fundamental properties of the system
- (2) Principles guiding design and evolution

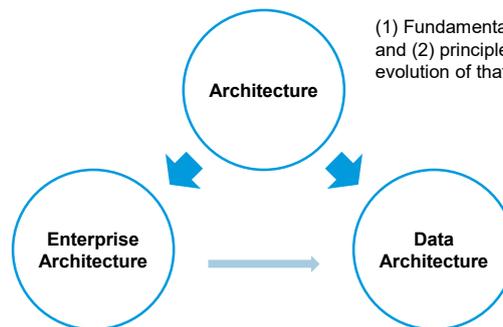


Architecture: scope

EA = architecture of the enterprise.

Key topics:

- What are the key building blocks of the enterprise?
- Capturing the big picture: insight, coherence, fitting the pieces of the puzzle together.



(1) Fundamental properties of a system, and (2) principles for guiding design and evolution of that system.

Architecture discipline at the intersection of enterprise architecture and DM.

Data architecture = architecture of the data landscape.

Key topics:

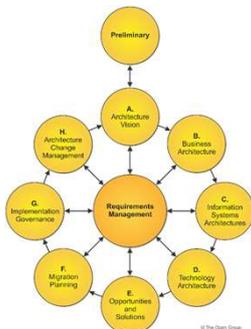
- Data needs
- Blueprints to realize those needs.



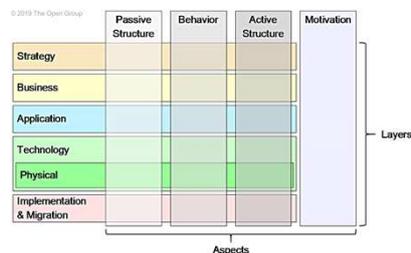
Frameworks

| | What | How | Where | Who | When | Why | |
|---------------------|--------------------------|------------------------|-----------------------------|-------------------------------|-----------------------|---------------------------|-----------------------|
| Enterprise | Inventory Identification | Process Identification | Distribution Identification | Responsibility Identification | Timing Identification | Motivation Identification | Strategic Context |
| Business Management | Inventory Definition | Process Definition | Distribution Definition | Responsibility Definition | Timing Definition | Motivation Definition | Business Context |
| Architecture | Inventory Representation | Process Representation | Distribution Representation | Responsibility Representation | Timing Representation | Motivation Representation | System Logic |
| Engineer | Inventory Specification | Process Specification | Distribution Specification | Responsibility Specification | Timing Specification | Motivation Specification | Technology Papers |
| Technician | Inventory Configuration | Process Configuration | Distribution Configuration | Responsibility Configuration | Timing Configuration | Motivation Configuration | Tool Conventions |
| Enterprise | Inventory Instantiation | Process Instantiation | Distribution Instantiation | Responsibility Instantiation | Timing Instantiation | Motivation Instantiation | Operational Instances |
| | Inventory Data | Process Data | Distribution Networks | Responsibility Assignments | Timing Cycles | Motivation Intentions | |

Zachman
Integral description. The “mother of all frameworks”.



TOGAF®
Open Group standard with a focus on the process of architecture development.



ArchiMate®
Open Group standard with a focus on modeling.



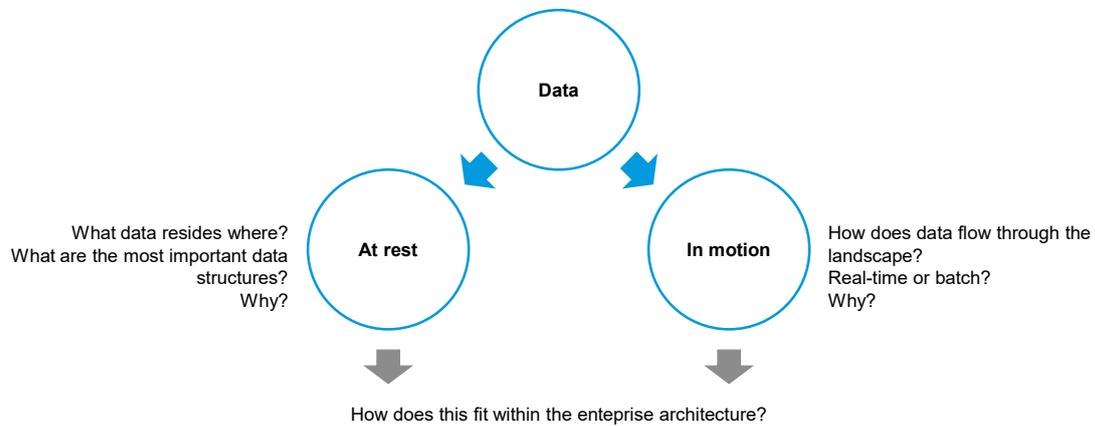
Enterprise data model



- ➔ **Conceptual data model:**
Most important business concepts/terms: what are the things we talk about in the organization?
- ➔ **Subject areas:**
Partitioning the data of the enterprise into logical groups (subject areas) of data about the same “thing”.
- ➔ **Logical data model:**
Structuring data for storage in a database management system in a platform-independent manner.
- ➔ **Physical data model:**
Model the way data in the database running on a database management system is physically stored.



Data in motion, data at rest



Exercise: data governance & data architecture

Setting:

An important task for the data architect is to prepare the organization for rapid development of products and services and the data associated with them. This helps the organization to seize opportunities and be ready to leverage new technologies effectively to achieve business success.

You work for an international IT training provider. This organization used to focus solely on the Dutch market but has recently taken over several companies across and even outside Europe. The organization wants to offer standardized training services internationally. The structure of the organization is being reshaped. Every country has its own headquarters and is responsible for the end-to-end process: registration, teaching, certification.

Assignment:

- 1) Describe the role of data architecture in this organization. Explain how it relates to the role of enterprise architecture. What are key choices in the enterprise architecture, and how can the data architect leverage them for his own work?
- 2) What key choices would you make around data governance? Central or decentralized data ownership? Why?

This assignment is done in groups of two to four persons. Document the outcomes of your discussion. These will be presented in the main session.



Case: Data management at a supermarket chain - DA



Context:

- The first steps on the Data Management Roadmap have been taken, and a hybrid operating model is chosen.
- The next step on the map is to get more value out of data.
- The organization has therefore decided to invest in an architecture capability.



Setting:

- The data team sets the following requirements:
 - We want to get a high-level overview of the interplay between processes, data, and systems for apples.
 - We want to decide on key issues around standardization, integration, flexibility, etc.
- The main processes are purchasing, marketing and sales.



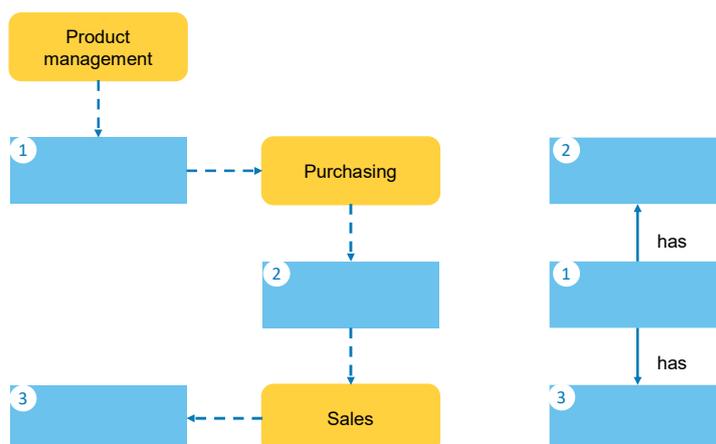
Challenge:

- We need a new architecture to sell our apples. Which labels should we use for (1), (2), and (3) as shown on the following slide?



Case: Data management at a supermarket chain - DA

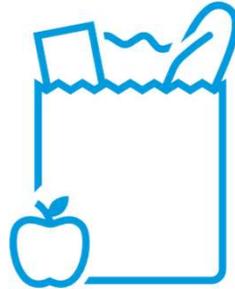
- The yellow boxes represent business functions. The blue boxes represent data inputs and outputs.
- Hint: Left is the process model. Right is a part of the conceptual data model. The labels should be the same.



Case: Data management at a supermarket chain - DA

Enterprise architecture

- How to deal with franchises vs owned stores.
- Where do we standardize processes? Where do we integrate processes through standardized data?
- Where do we need flexibility and where should we optimize for performance?
- What is the impact of off-the-shelf systems versus home-grown systems on processes and data?



Data architecture

Given the choices that were made at the enterprise architecture level:

- What are the main groups of data (subject areas, data clusters)?
- What are their properties (stable, frequency of changing structure, frequency of updates)?
- What are the key data flows and how do we want to deal with them?
- What does that mean for data structures?



Practice questions

1. Which statement is NOT true about the enterprise-wide data model?

- A. The corporate data architect owns the enterprise-wide data model.
- B. The enterprise-wide data model is driven by the business.
- C. Subject areas are areas of concern for the corporation.
- D. The enterprise data model will frequently change.

2. Which Enterprise Architecture Framework defines artifacts in a 6 x 6 matrix, with interrogatives (what, how, where, etc.) as columns and stakeholder perspectives (executive, business, architect, etc.) as rows?

- A. TOGAF.
- B. FEAR.
- C. Zachman.
- D. Kimball.
- E. ArchiMate.



CHECK-IN

HOW DO YOU LOOK AT DATA MODELING AND DESIGN?

Mandatory "work" _____

IT _____

Solo _____

Super cool

Business

Group

_____ Super cool

_____ Business

_____ Group

© Van Haren Publishing
39

DATA MODELING & DESIGN

DATA MODELING IS THE PROCESS OF DISCOVERING, ANALYZING AND SCOPING DATA REQUIREMENTS; TO SUBSEQUENTLY REPURPOSE AND COMMUNICATE THESE IN THE FORM OF A DATA MODEL.

Data Modeling and Design

Definition: Data modeling is the process of discovering, analyzing and scoping data requirements, and then representing and communicating these data requirements in a precise form called the data model. This process is iterative and may include a conceptual, logical, and physical model.

Goal: 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.

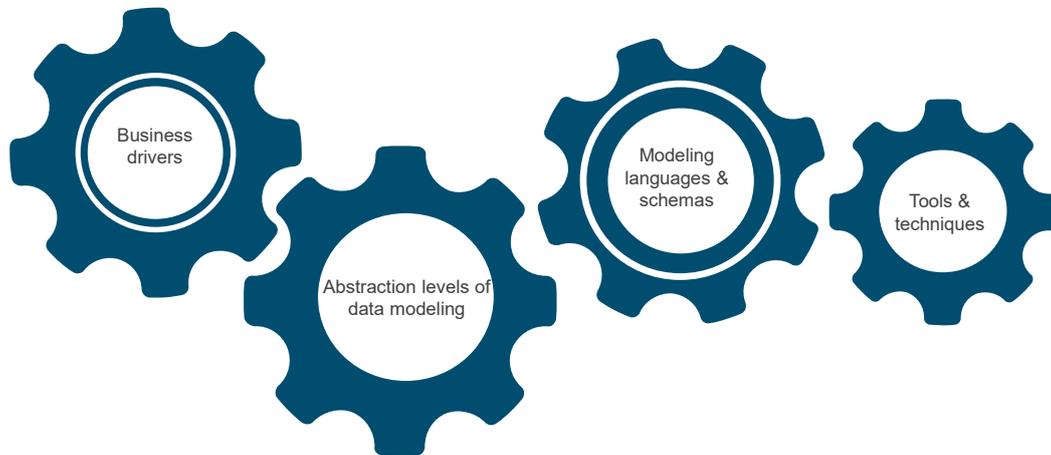
| | | |
|---|--|--|
| <p>Inputs:</p> <ul style="list-style-type: none"> Existing data models and databases Data standards Data sets Initial data requirements Original data requirements Data architecture Enterprise taxonomy | <p>Activities:</p> <ol style="list-style-type: none"> Plan for Data Modeling (P) Build the Data Models (D) <ol style="list-style-type: none"> Create the Conceptual Data Model Create the Logical Data Model Create the Physical Data Model Review the Data Models (C) Manage the Data Models (O) | <p>Deliverables:</p> <ul style="list-style-type: none"> Conceptual Data Model Logical Data Model Physical Data Model |
| <p>Suppliers:</p> <ul style="list-style-type: none"> Business Professionals Business Analysts Data Architects Database Administrators and Developers Subject Matter Experts Data Stewards Metadata Administrators | <p>Participants:</p> <ul style="list-style-type: none"> Business Analysts Data Modelers | <p>Consumers:</p> <ul style="list-style-type: none"> Business Analysts Data Modelers Database Administrators and Developers Software Developers Data Stewards Data Quality Analysts Data Consumers |
| <p>Techniques:</p> <ul style="list-style-type: none"> Naming conventions Database design Database type selection | <p>Tools:</p> <ul style="list-style-type: none"> Data modeling tools Lineage tools Metadata repositories Data model patterns Industry data models | <p>Metrics:</p> <ul style="list-style-type: none"> Data model validation measurement |

(P) Planning, (C) Control, (D) Development, (O) Operations

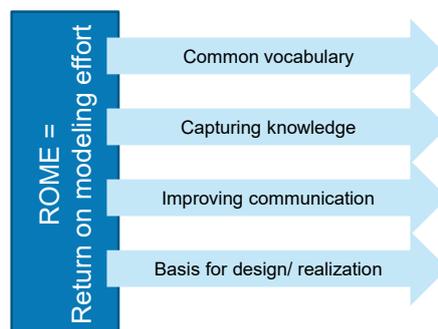
Figure taken from: DAMA DMBOK v2

© Van Haren Publishing
40

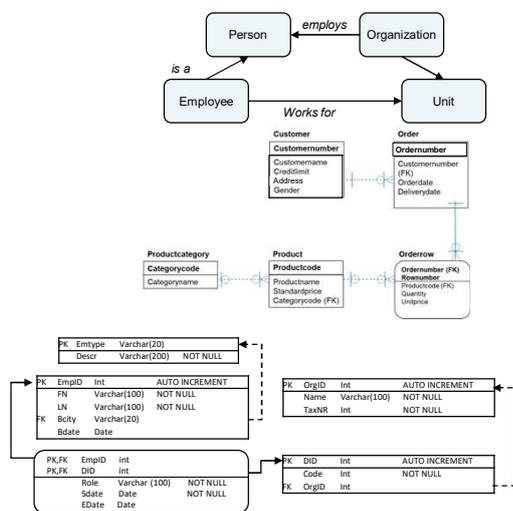
Topics



Business drivers



Abstraction ess



Conceptual

- Focus: understanding concepts and relations between relations.
- Languages: 'informal', FBM, ERD.

Logical

- Focus: data structures:
 - o Characteristics defined
 - o Primary and foreign key identified
 - o Technology independent
- Languages: ERD, UML.

Physical

- Focus: internal data storage in database systems
- Languages : ERD.

Theory: schema types

Different goals lead to different schema types.

Goals: a) understand domain, b) design transaction system, c) design BI-system

Entity Relation Diagram

Primary for relation models.



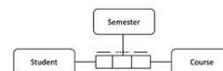
Object Oriented (UML)

Primary for software systems. The class diagram is widely used.



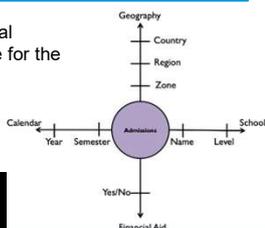
Fact based (ORM2, FCO/IM)

Has a strong connection with the business rules.



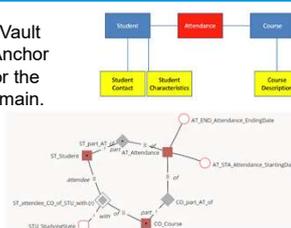
Dimensional

Dimensional models are for the BI-domain.



Time-based

Data Vault and Anchor are for the BI-domain.



NoSQL

No (standard) visual notation known. NoSQL is about other forms of databases:

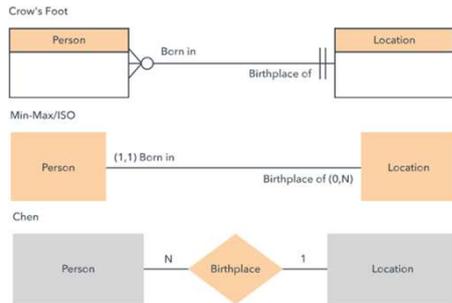
- Key/value
- Wide column
- Graph
- Etc.

The figures for the dimensional model, data vault model, and anchor model are taken from DAMA DMBOK v2.

Languages & schemas

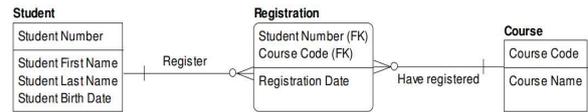
ERD notation

Alternatives for cardinality.



ERD notation

Example of a logical data model.

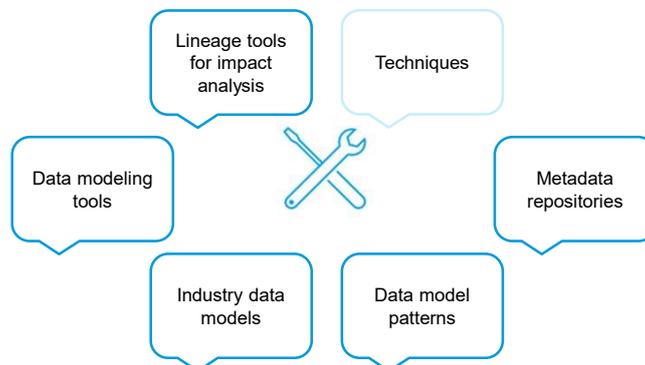


Points of interest:

- Entity types
- Attribute types + keys
- Relationship types + name
- Cardinality



Tools and techniques



Case: Data management at a supermarket chain - DM&D



Context:

- For the supermarket case, work is on-going.
- The Data Governance board is meeting frequently.
- Responsibilities have shifted from just a group of enthusiasts to a more organization-wide movement with not only apples, but more and more business owners involved.
- The business executive is pleased that he has received a high-level data architecture.



Setting:

- Based on the architecture analysis, a decision was made to consolidate several systems used for purchasing.
- For the 'owned' stores, the decision can be made and communicated in a top-down fashion.
- For the franchises, this requires more negotiation.
- An additional modeling specialist is added to the internal team.
- The business executive has asked the working team to prepare for the upcoming consolidation.

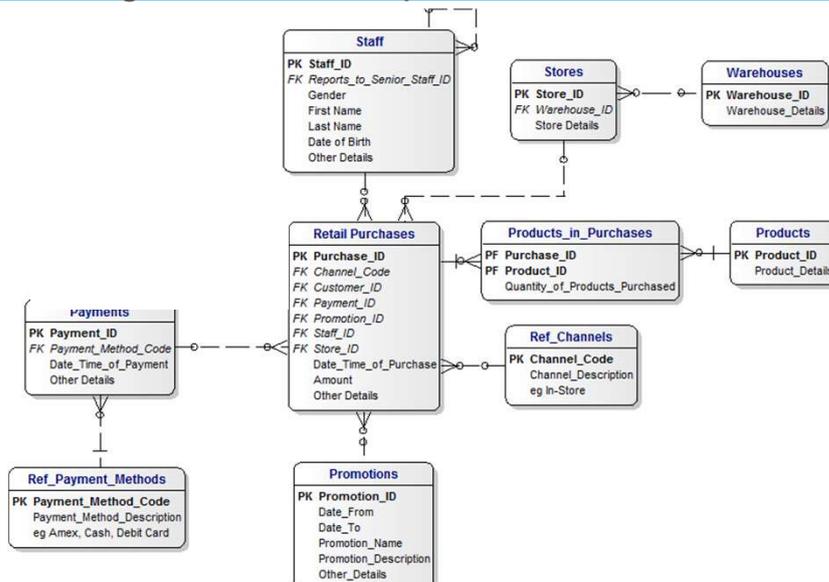


Challenge:

- You are going to have a chat with a modeling specialist with a lot of experience in data modeling for the supermarket industry using ERD.
- Where in the following data model should the entity (block) 'customers' be added?



Case: Data management at a supermarket chain - DM&D



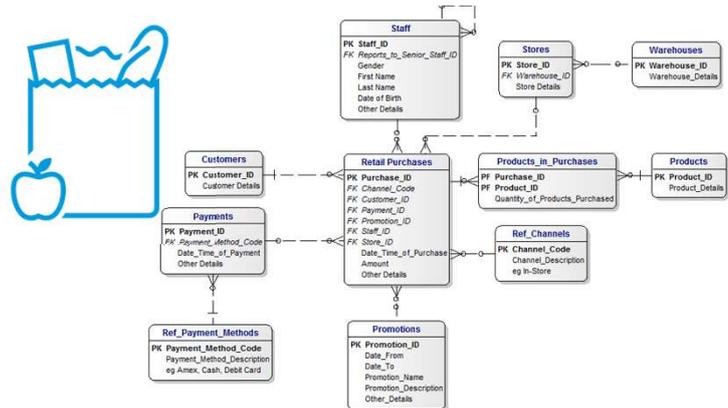
Case: Data management at a supermarket chain - DM&D

Using tools and reference models

- The business model for supermarkets is fairly standard. We know which terms/concepts are used frequently. Vendors tend to have good reference models that we can reuse.
- In a merger/acquisition situation, data profiling can be used to assess whether the data of one party can be used in the systems of another party.

Logical data model

http://www.databaseanswers.org/data_models/enterprise_data_model_for_retail/index.htm



Practice questions

1. When a data modeler would like to roll back a change to a data model, which function would they use?

- Change Control.
- Model Merge.
- Versioning.
- Sub-modeling.

2. Which is the highest level of these data model types?

- Operating Model.
- Conceptual Model.
- Logical Model.
- Physical Model.
- Super Model.

