COURSEWARE

DATA QUALITY MANAGEMENT

BASED ON CDMP

Van Haren PUBLISHING

MICHEL DEKKER

Data Quality Management
Based on CDMP

Colophon

Title:	Data Quality Management
Subtitle:	Based on CDMP
Authors:	Michel Dekker
Publisher:	Van Haren Publishing, ‘s-Hertogenbosch
ISBN Hard Copy:	978 94 018 1354 9
Edition:	First edition, first print, September, 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

No part of this publication may be reproduced in any form by print, photo print, micro-film 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 in-completeness in this publication.

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, frameworks and standards within four domains:
- IT Management
- Architecture (Enterprise and IT)
- Business Management and
- Project Management

Van Haren Publishing is publishing on behalf of leading organizations and companies: Agile Consortium, World Commerce and Contracting, IAOP, IPMA World, KNVI, PMI-NL, NLAIC and The Open Group.

Van Haren Publishing is part of the Van Haren Group and additional to the book publishing also provides the following services: accredited training materials and e-learning through Van Haren Learning Solutions, as well as independent professional certification via examination through Van Haren Certify.

Topics are (per domain):

IT Management	IT Service Management	FitSM, ISM®, ISO/IEC20000, IT4IT®, ITIL®, VerISM®, SAF, TRIM, XLA®
	Data Management	Data literacy, Data visualization, DMBOK
	IT Asset Management	HAM, ITAM, SAM
	IT Security Management	BIO, ISO/IEC27001, NIS2
	Test Management	CTAP
	Application Management	ASL
	Other	eCF, IT-CMF, Scrum
Project Management	Project Management	Half Double, ICB, ISO/IEC21500, P3.express, PM2, PMBOK Guide, Praxis, PRINCE2
	Agile	Agile, Agile PM
	Other	PMO
Business Management	Operations Management	Lean, Lean Six Sigma, OBM, OMC, RASCI
	Contract Management	CATS CM, CATS RVM, IACCM World
	Business Information Management	BiSL, DID
	Artificial Intelligence	AI, Generative AI
	Outsourcing	OPBOK
Enterprise Architecture	Enterprise Architecture	BIAN, TOGAF
	Modeling	ArchiMate, BPMN
	Software Architecture	ISAQB
	Other	Open Agile Architecture

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

Table of Content

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.

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

(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.	

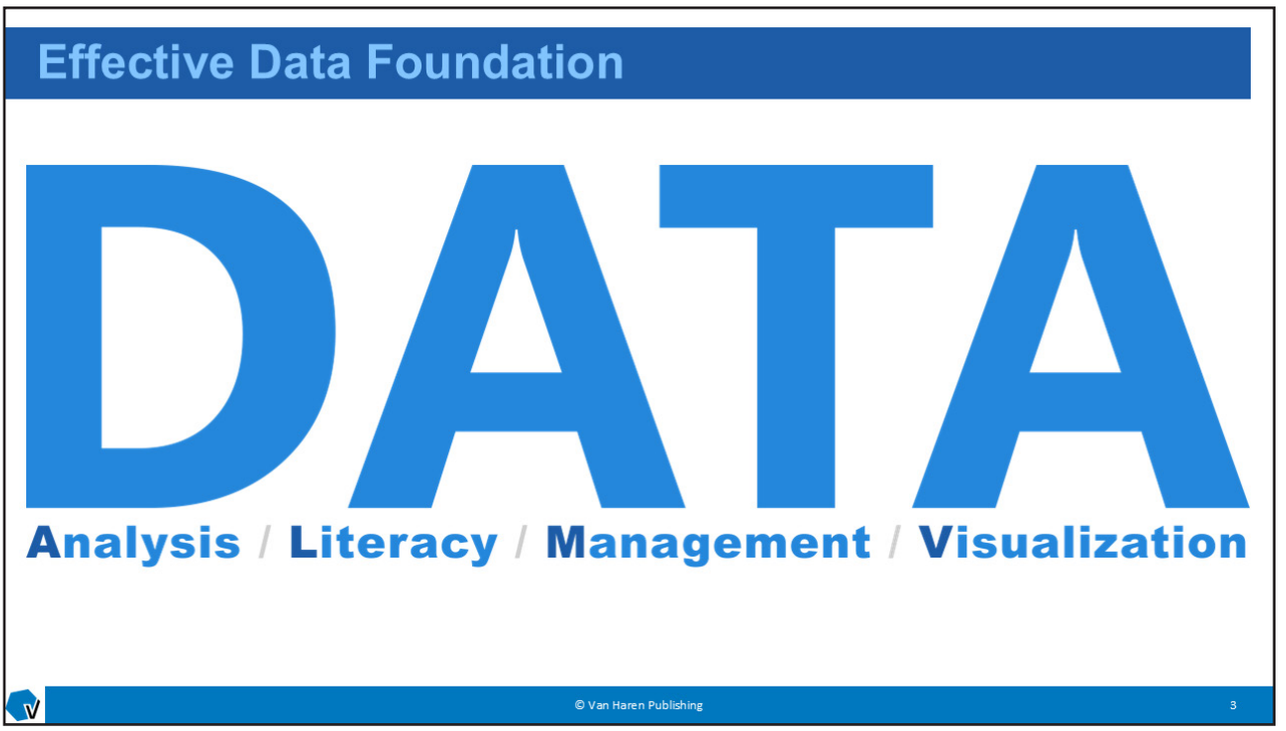
Data Quality Management

©2025 - All training materials are sole property of Van Haren Publishing BV
and are not to be reproduced in any form or shape without written permission.

Effective Data Foundation

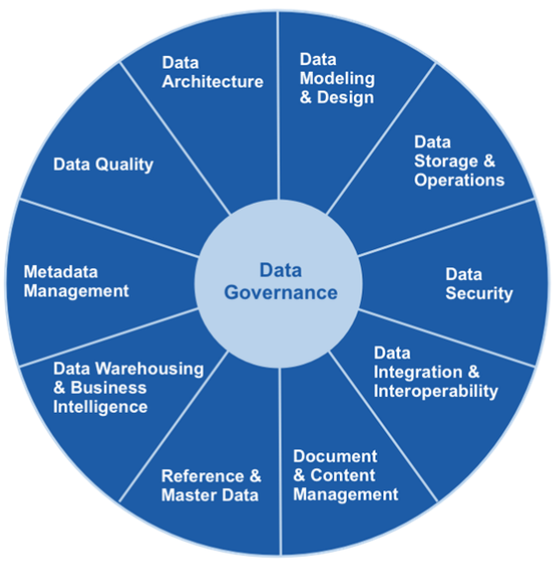
Not-for-profit collective, which enables **professionals** to **leverage data** to make **sustainable business decisions**

© Van Haren Publishing 2


Effective Data Foundation
DATA
Analysis / Literacy / Management / Visualization
© Van Haren Publishing
3


Data Management
Data management is about
people, processes,
and technology,
in that order.
Source: unknown
© Van Haren Publishing
4

DAMA Wheel



Source: Dama.org

© Van Haren Publishing 5

Certification

ASSOCIATE	PRACTITIONER	MASTER	FELLOW
CDMP Associate	CDMP Practitioner	CDMP Master	CDMP Fellow
Membership: Central	Membership: Central	Membership: Central	Membership: Central
Industry experience: 6 months > 5 years	Industry experience: 2 years - 10 years	Industry experience: Minimum 10 years^	Industry experience: 25 years plus
Requirements: 60% pass Data Management Fundamentals exam	Requirements: 70% pass in Data Management Fundamentals exam and 70% pass in 2 specialist exams	Requirements: 80% pass in Data Management Fundamentals exam and 80% pass in 2 specialist exams	Requirements: Globally recognised & respected thought leadership; Significant contribution to Data Management profession; CDMP Master Contribution to CDMP & DMBOK; By nomination

© Van Haren Publishing 6

Training Agenda

- Introduction
- Essential concepts
- Activities
- Tools & Techniques
- Implementation
- Governance

© Van Haren Publishing

What is Data?

© Van Haren Publishing

Data & Reality

© Van Haren Publishing 9

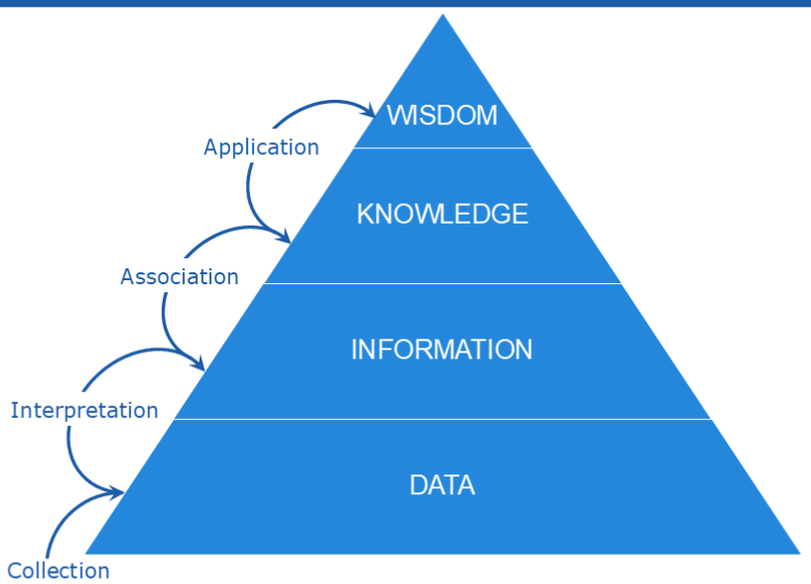
What is Data?

Describes a ***quality*** *or* ***quantity*** *of some* ***object*** *or* ***event****.*

Source: unknown

© Van Haren Publishing 10

Why do we use Data?



Source: Jones, Ben. Data Literacy Fundamentals: Understanding the Power & Value of Data

© Van Haren Publishing

First principle for understanding data

Source: Wheeler, Donald J. Understanding variation – the key to managing chaos

© Van Haren Publishing

Second principle for understanding data

While **every data set** contains **noise**, **some data sets** may contain **signals**.

Therefor, before you can **detect a signal** within any given data set, you must **first filter out the noise**.

Source: Wheeler, Donald J. Understanding variation – the key to managing chaos

© Van Haren Publishing 13

Data Quality

Data is of high **quality**, if the data is **fit for** the intended **purpose** of use

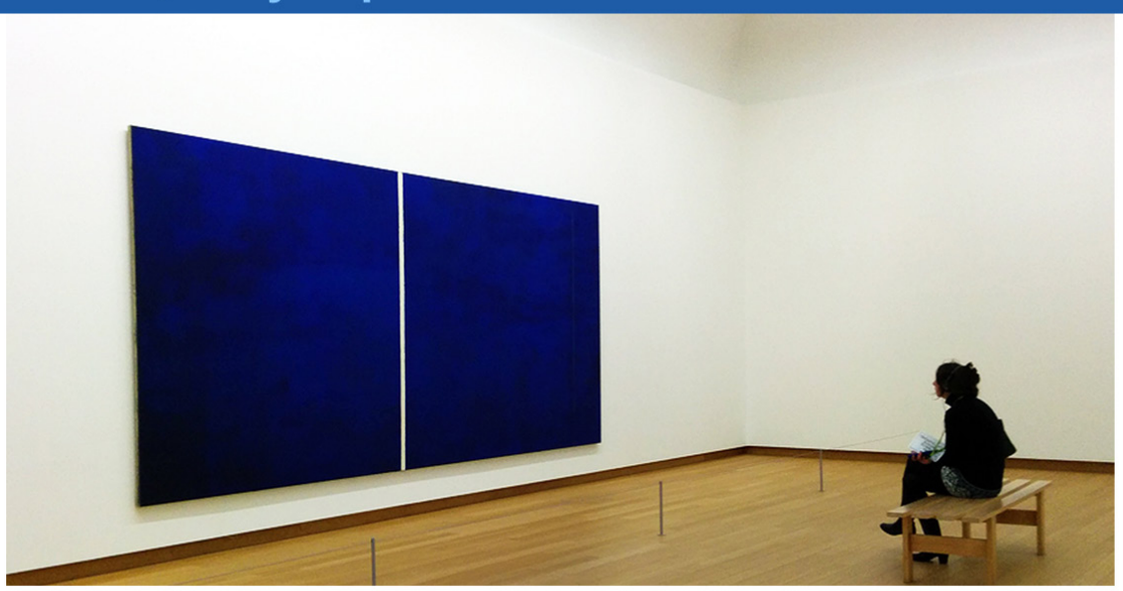
© Van Haren Publishing 14

Data Quality

Data is of high **quality**, if the data correctly **represents the real-world** construct that the data describes

© Van Haren Publishing 15

Data Quality is personal



© Van Haren Publishing 16

Data Quality is situational

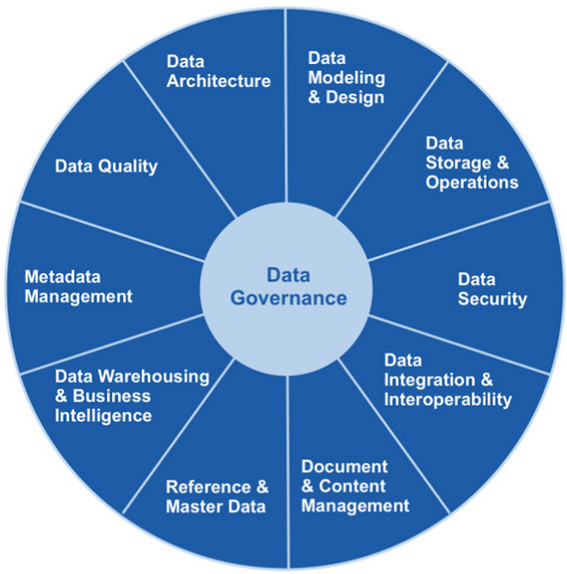
© Van Haren Publishing

Definition Data Quality Management

The planning, implementation, and control of activities that apply techniques for collecting and handling data ensuring it **addresses the needs** of the enterprise and local consumer and is **fit for use**.

© Van Haren Publishing

Relations with other knowledge areas



© Van Haren Publishing

Relation with Data Governance

Ownership

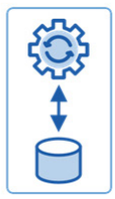
Stewardship

Resolving/prevent issues

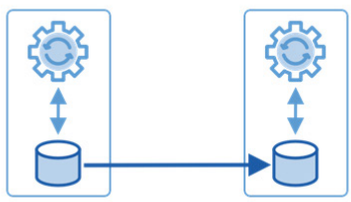
© Van Haren Publishing

Relation with Data Architecture

Data at rest

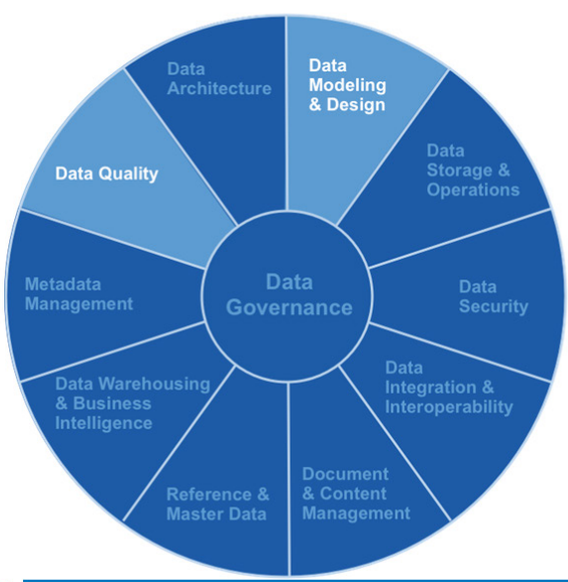


Data in motion



© Van Haren Publishing 21

Relation with Data Modeling & Design



Keys (uniqueness)

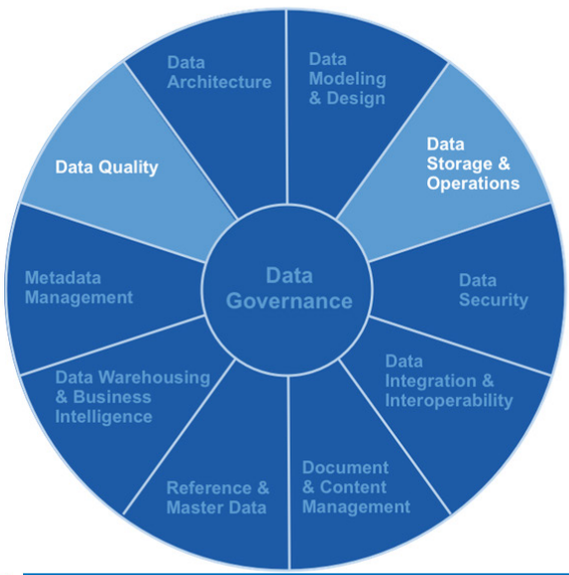
Referential integrity

Normalization

Include DQ values

© Van Haren Publishing 22

Relation with Data Storage & Operations



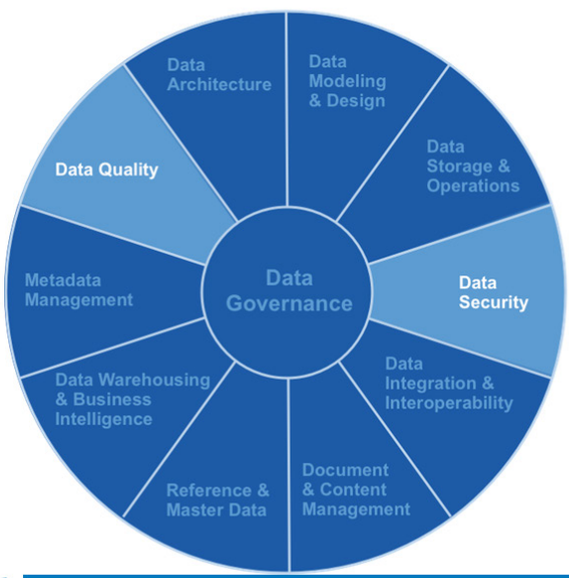
Operations vs Quality

Legacy

Assurance vs Scalability

© Van Haren Publishing

Relation with Data Security



Integrity

Encryption/masking

CRUD

© Van Haren Publishing

Relation with Data Integration & Interoperability

Keep the metadata (linked)

Test and reconcile

Cause of DQ issues

© Van Haren Publishing

Relation with Document & Content Management

Unstructured data

Content vs Record

© Van Haren Publishing

Relation with Reference & Master Data

Critical data

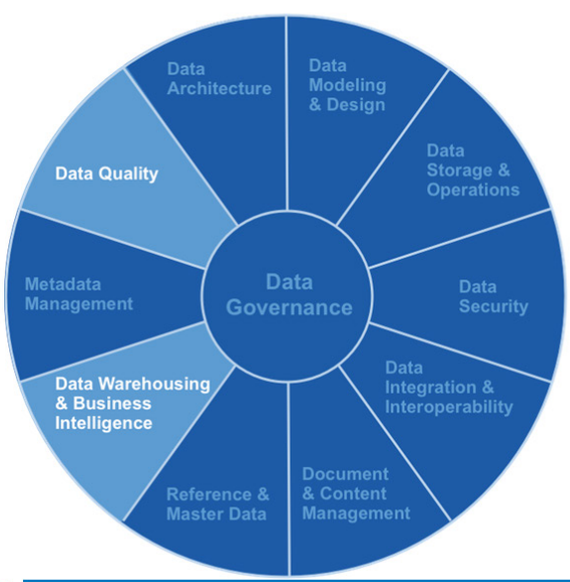
Domain values

Source for rule enforcement

Change management

© Van Haren Publishing

Relation with Data Warehousing & Business Intelligence



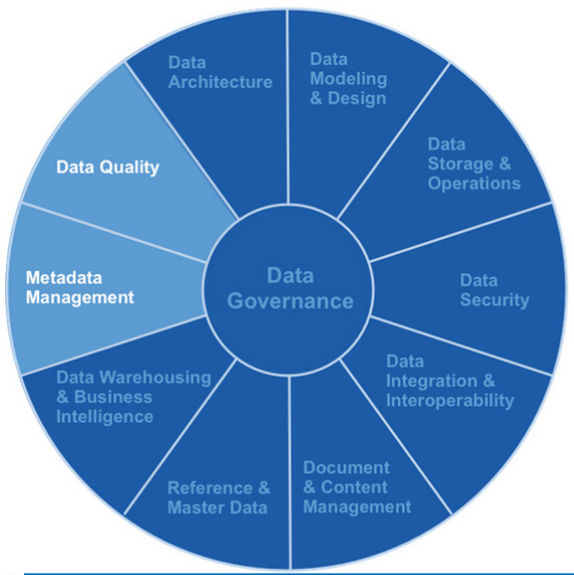
End of the value stream

Aggregate data

Changes over time

© Van Haren Publishing

Relation with Metadata Management



Data Quality is about meeting expectations.

Metadata is a primary means of clarifying expectations.

© Van Haren Publishing

Business Drivers

- Enhancing stakeholders **experience** and organizations's **reputation**
- Increase the **effectiveness** of the organization
- Reducing **risks** and **costs** associated with poor quality data
- Improving organizational **efficiency** and **productivity**

© Van Haren Publishing

Goals

Make data **fit for purpose** based on data consumers' requirements

Defining **standards** and **specifications** for Data Quality controls as part of the data lifecycle

Defining and implementing 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

© Van Haren Publishing

Principles

Criticality

Embedded in business processes

Standards-driven

Objective measurement

Prevention

Transparency

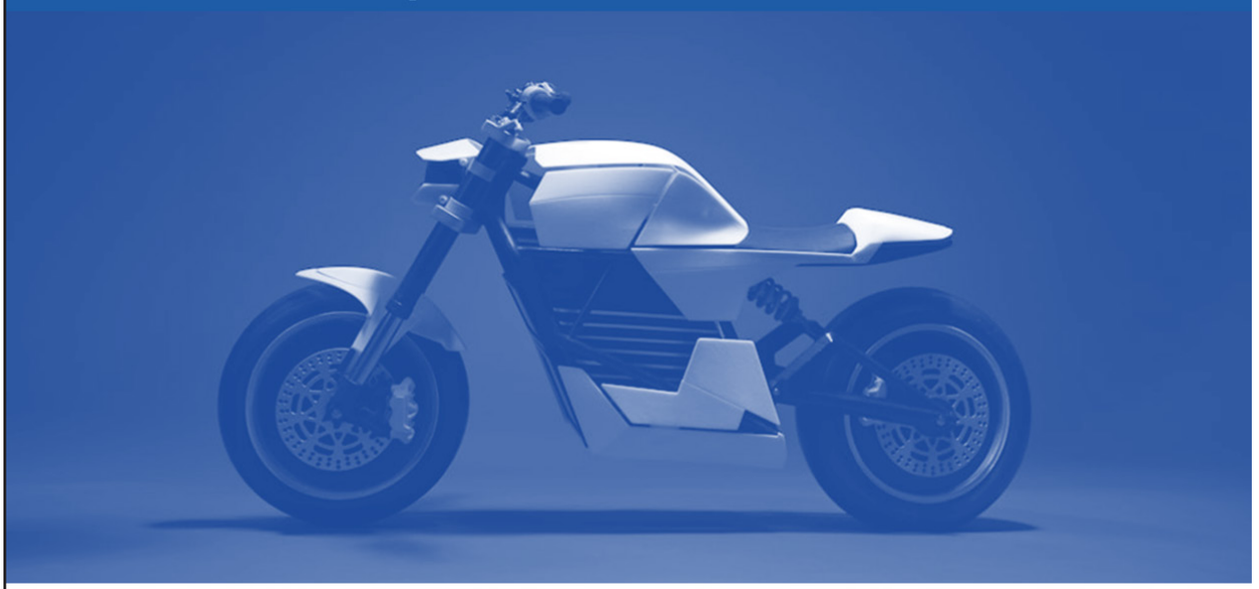
Connected to service levels

Root cause remediation

Systematically enforced

© Van Haren Publishing

Essential Concepts



© Van Haren Publishing 33

First principle of Data Quality Management

*Focus improvement efforts on **data** that is **most important** to the organization and its customers.*

Source: DAMA DMBoK

© Van Haren Publishing 34

Critical data

- (Regulatory) compliance
- Business operational needs
- Measuring product quality and customer satisfaction
- Business strategy

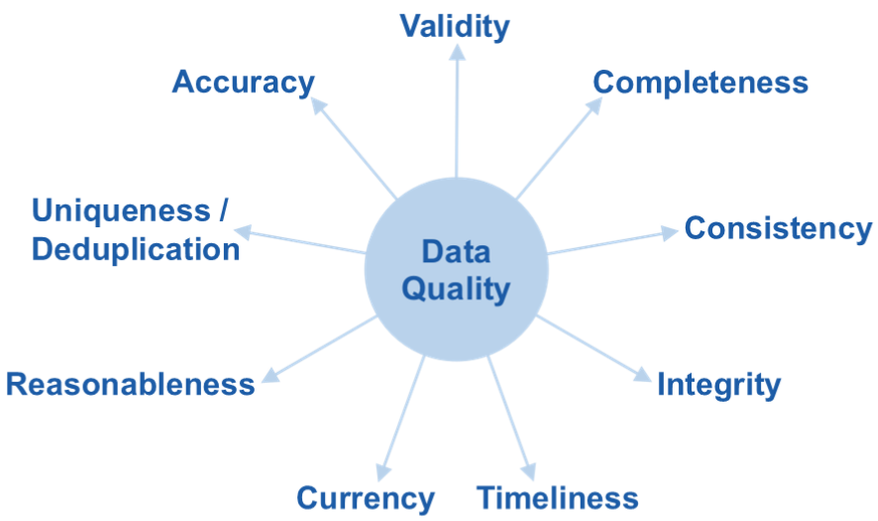
© Van Haren Publishing

Data Quality Dimensions

a measurable feature or characteristic of data

© Van Haren Publishing

Data Quality Dimensions



© Van Haren Publishing 37

Dimension - Validity

whether data values are **consistent** with a **defined domain** of **values**

© Van Haren Publishing 38

Dimension - Completeness

whether **all** the **required** data is **present**

© Van Haren Publishing 39

Dimension - Consistency

ensuring that data values are **coded** using the **same approach**, **assessment** and **valuation** criteria

© Van Haren Publishing 40

Dimension - Integrity

the **lack of incoherent** values and **broken relationships** in data

© Van Haren Publishing 41

Dimension - Timeliness

the **time** that data needs to become **accessible** to a **user** after its **capture** or **update**

© Van Haren Publishing 42

Dimension - Currency

the date that the data was **last updated** relative to now and the **likelihood** that it is still **correct**

© Van Haren Publishing 43

Dimension - Reasonableness

whether a data **pattern meets expectations**

© Van Haren Publishing 44

Dimension – Uniqueness / Deduplication

no real-world entity exists **more than once** within the data set

© Van Haren Publishing 45

Dimension – Accuracy

the degree that data **correctly represents** 'real-life' entities

© Van Haren Publishing 46

Data Quality Dimensions

Dimension	Description
Validity	whether data values are consistent with a defined domain of values
Completeness	whether all the required data is present
Consistency	ensuring that data values are coded using the same approach, assessment and valuation criteria
Integrity	the lack of incoherent values and broken relationships in data
Timeliness	the time that data needs to become accessible to a user after its capture or update
Currency	the date that the data was last updated relative to now and the likelihood that it is still correct
Reasonableness	whether a data pattern meets expectations
Uniqueness/Deduplication	no real-world entity exists more than once within the data set
Accuracy	the degree that data correctly represents 'real-life' entities

© Van Haren Publishing

Which DQ dimension fits best?

1. The value in the "Customer age" field is "11001"

a. Validity
b. Completeness
c. Consistency
d. Integrity
e. Timeliness
f. Currency
g. Reasonableness
h. Uniqueness/Deduplication
i. Accuracy

© Van Haren Publishing

Which DQ dimension fits best?

2. Each student must have a date of birth

a. Validity
b. Completeness
c. Consistency
d. Integrity
e. Timeliness
f. Currency
g. Reasonableness
h. Uniqueness/Deduplication
i. Accuracy

© Van Haren Publishing 49

Which DQ dimension fits best?

3. All sales transactions have a timestamp

a. Validity
b. Completeness
c. Consistency
d. Integrity
e. Timeliness
f. Currency
g. Reasonableness
h. Uniqueness/Deduplication
i. Accuracy

© Van Haren Publishing 50

Which DQ dimension fits best?

4. Employee A354 from Germany is 12 years old and has a driving license (car)

a. Validity
b. Completeness
c. Consistency
d. Integrity
e. Timeliness
f. Currency
g. Reasonableness
h. Uniqueness/Deduplication
i. Accuracy

© Van Haren Publishing

Which DQ dimension fits best?

5. Citizen D345 is registered as married to K693. Citizen K693 is registered as being single.

a. Validity
b. Completeness
c. Consistency
d. Integrity
e. Timeliness
f. Currency
g. Reasonableness
h. Uniqueness/Deduplication
i. Accuracy

© Van Haren Publishing

Which DQ dimension fits best?

6. Our total sales in the month of January is 500 according to our ERP system. For this same period our Data Warehouse shows a total sales of 935

a. Validity
b. Completeness
c. Consistency
d. Integrity
e. Timeliness
f. Currency
g. Reasonableness
h. Uniqueness/Deduplication
i. Accuracy

© Van Haren Publishing 53

Which DQ dimension fits best?

7. 85% of all employee records have “null” as a value for the field “Reports to”

a. Validity
b. Completeness
c. Consistency
d. Integrity
e. Timeliness
f. Currency
g. Reasonableness
h. Uniqueness/Deduplication
i. Accuracy

© Van Haren Publishing 54