COURSEWARE

REFERENCE AND MASTER DATA MANAGEMENT

BASED ON CDMP

Van Haren PUBLISHING

MICHEL DEKKER

Reference and Master Data Management
Based on CDMP

Colophon

Title:	Reference and Master Data Management
Subtitle:	Based on CDMP
Authors:	Michel Dekker
Publisher:	Van Haren Publishing, ‘s-Hertogenbosch
ISBN Hard Copy:	978 94 018 1396 9
Edition:	First edition, first print, November, 2025
Design:	Van Haren Publishing, ‘s-Hertogenbosch

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

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.	


Reference & Master Data Management
COURSEWARE
©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.

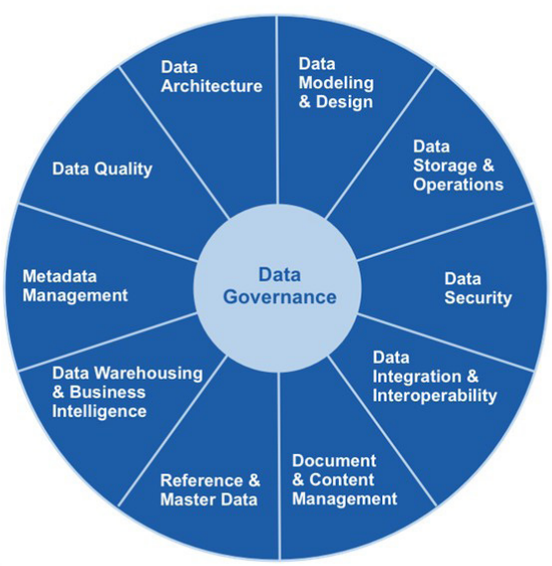

Effective Data Foundation
Not-for-profit collective, which enables professionals to leverage data to make sustainable business decisions
© Van Haren Publishing
2

Effective Data Foundation

DATA

Analysis / Literacy / Management / Visualization

DAMA Wheel



Reference and Master Data

Definition: Managing reconciled and integrated data through stewardship and semantic consistency in support of enterprise-wide needs to share its data assets.

Goals:
1. Ensuring the organization has complete, consistent, current, authoritative Master and Reference Data across organizational processes.
2. Enabling Master and Reference Data to be shared across enterprise functions and applications.
3. Lowering the cost and reducing the complexity of data usage and integration through standards, common data models, and integration patterns.

Business Drivers

Inputs:	Activities:	Deliverables:
• Candidate Data Stores and Values • Cross Functional Requirements • Industry Standards • Metadata • Purchased Data and/or Open Data and Code Sets • Business Rules	1. Define Drivers and Requirements (P) 2. Evaluate and Assess Data Sources (P) 3. Define Architectural Approach (D) 4. Model Data (D) 5. Define Stewardship and Maintenance Processes (C) 6. Establish Governance Policies (C) 7. Implement Data Sharing / Integration Services (D,O) 1. Acquire Data Sources for Sharing 2. Publish Reference and Master Data	• Master and Reference Data Requirements • Data Models and Integration Patterns • Reliable Reference and Master Data • Reusable Data Services • Data Exception Reports

Suppliers:	Participants:	Consumers:
• Subject Matter Experts • Data Stewards • Application Developers • Data Providers • Business Analysts • Infrastructure Systems Analysts	• Data Analysts • Data Modelers • Data Stewards • Data Integrators • Data Architects • Data Quality Analysts	• Subject Matter Experts • Data Integrators • Application Users • Application Developers • Solution Architects

Technical Drivers

Techniques:	Tools:	Metrics:
• Conditions-of-use agreements • Business key cross references • Processing Log analysis	• Data Integration Tools • Data Remediation Tools • Operational Data Stores • Data Sharing Hubs • Data Modeling Tools • Metadata Repositories • MDM Application Platforms	• Data Quality and Compliance • Data Change Activity • Data Ingestion and Consumption • Data Sharing Availability • Data Steward Coverage • Data Sharing Volume and Usage

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

Source: Dama.org

Certification

ASSOCIATE	PRACTITIONER	MASTER
Required experience none	Required experience 2+ years	Required experience 10+ years
Exam requirements 60% pass Data Management Fundamentals exam	Exam requirements 70% pass Data Management Fundamentals exam + 2 Specialist exams	Exam requirements 80% pass Data Management Fundamentals exam + 2 Specialist exams

Source: DAMA

5

Training Agenda

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

6

Introduction

*You can't **manage** what you don't **define**, and you can't **trust** what you don't **master**.*

Anonymous (Data Governance proverb)

 7

What is Data?

 8

Data & Reality

 9

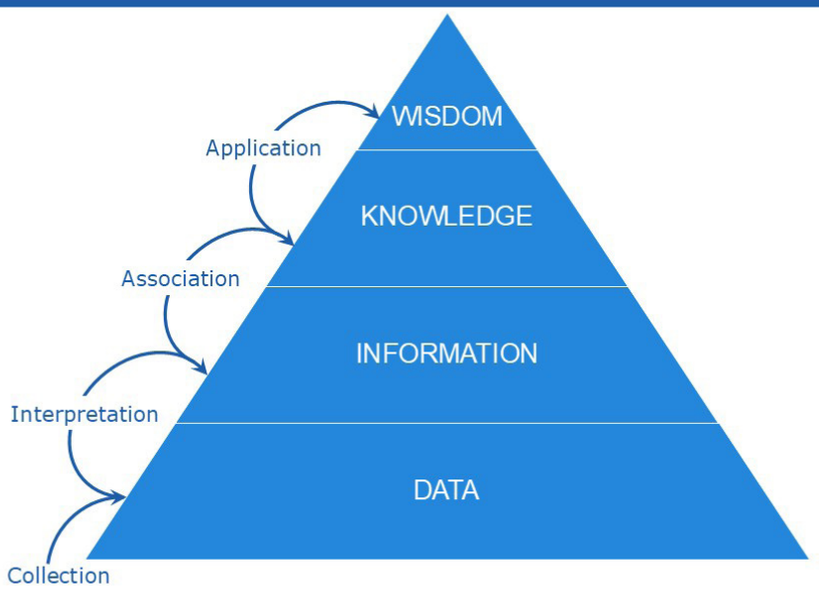
What is Data?

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

Source: unknown

 10

Why do we use Data?



Source: Jones, Ben. Data Literacy Fundamentals: Understanding the Power & Value of Data © Van Haren Publishing 11

First principle for understanding data

Source: Wheeler, Donald J. Understanding variation – the key to managing chaos © Van Haren Publishing 12

Definition

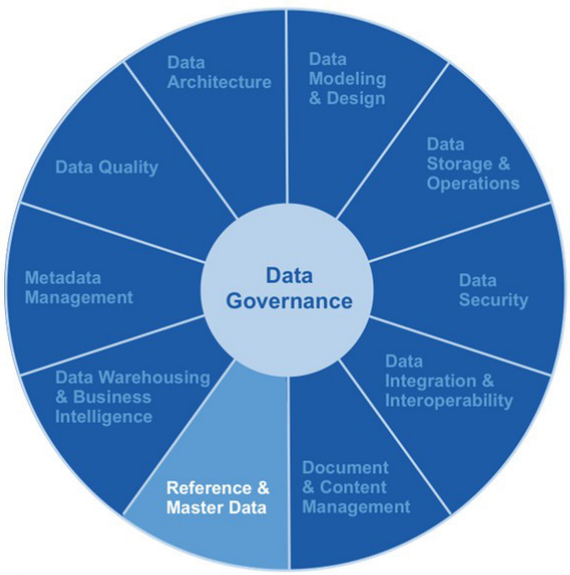
Managing **reconciled** and **integrated** data through **stewardship** and **semantic consistency** in support of enterprise-wide needs to **share its data assets**.

 13

Relations with other knowledge areas

 14

Relation with Data Governance

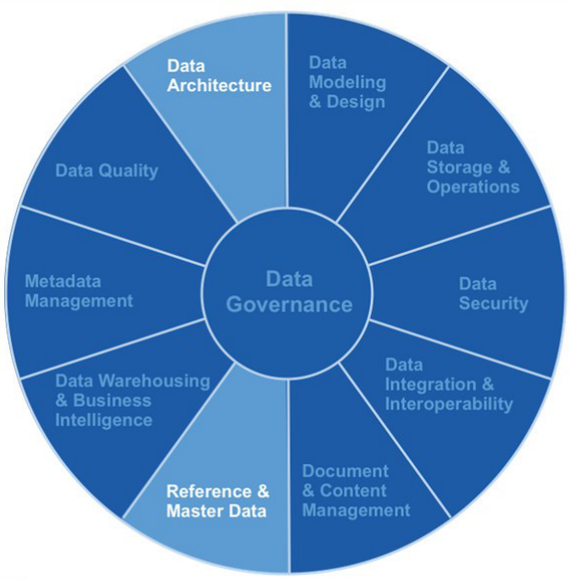


WHY & WHO

vs.

WHAT & HOW

Relation with Data Architecture

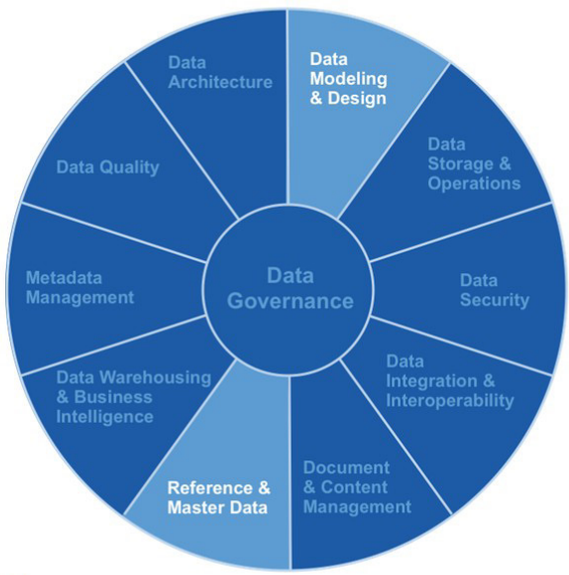


Architecture **designs** the house,

and

RMDM **defines** the **shared objects** that live inside it.

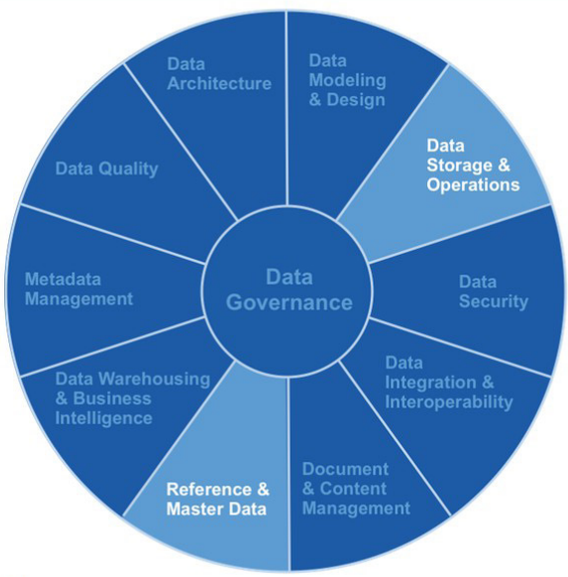
Relation with Data Modeling & Design



Data Modeling **defines structure**

RMDM **enforces consistency** and **control**

Relation with Data Storage & Operations



DS&O defines **WHERE** & **HOW** data lives

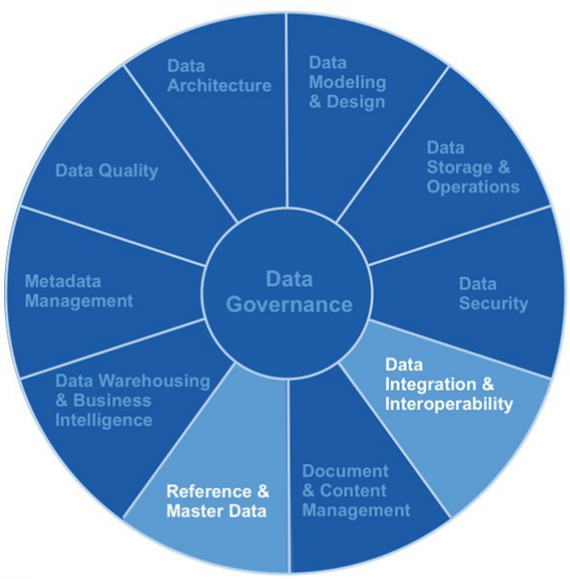
RMDM defines **WHAT** data must stay consistent and shared

Relation with Data Security

Data Security defines **protection** and **control**

RMDM defines **what** must be protected and by **whom**

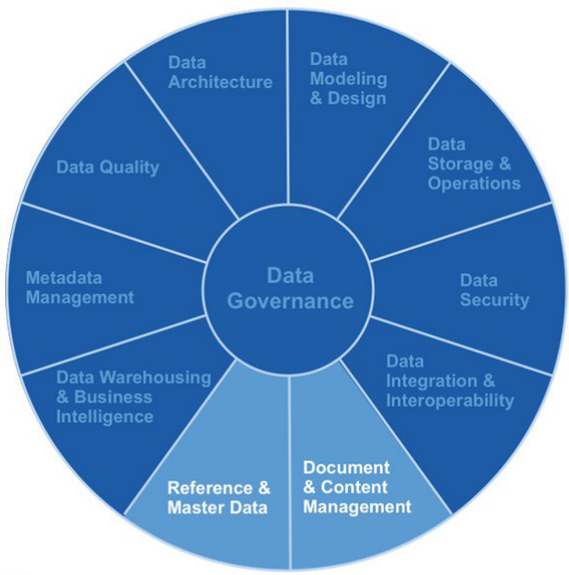
Relation with Data Integration & Interoperability



Data Integration & Interoperability defines **how data flows**

RMDM defines **what flows**.

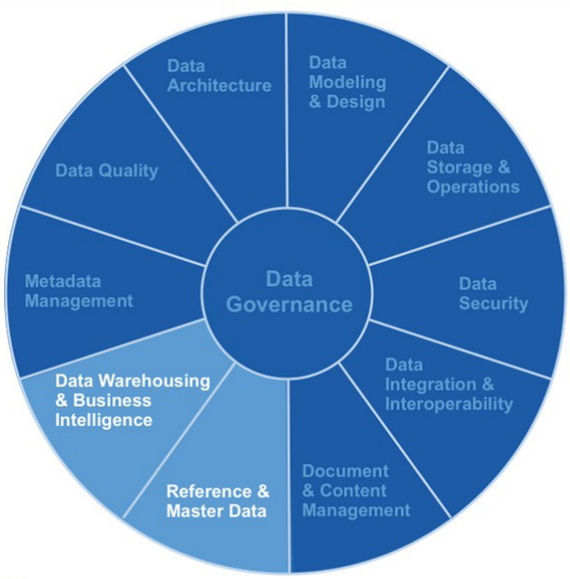
Relation with Document & Content Management



Document & Content Management organizes **unstructured** information.

RMDM provides the structure that gives that content **meaning**.

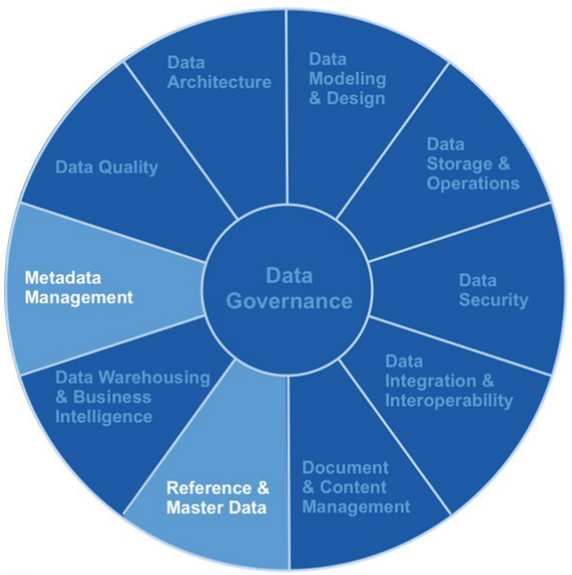
Relation with Data Warehousing & Business Intelligence



Data Warehousing & BI delivers **visibility**

RMDM ensures **credibility**

Relation with Metadata Management



Metadata Management defines **context** and **meaning**

RMDM provides the **authoritative data** those definitions describe

Relation with Data Quality

Data Quality ensures data is correct and **fit for purpose**

RMDM ensures that correct data is **shared and sustained**

Business Drivers

Meeting organizational data requirements

Managing the costs of data integration

Managing Data Quality

Reducing risk

Goals & Principles

Ensure **complete**, **consistent**, **current**, and **authoritative** data

Enable data **sharing** across enterprise functions and applications

Lower the **cost** and **complexity** of data usage and integration

Shared data

Ownership

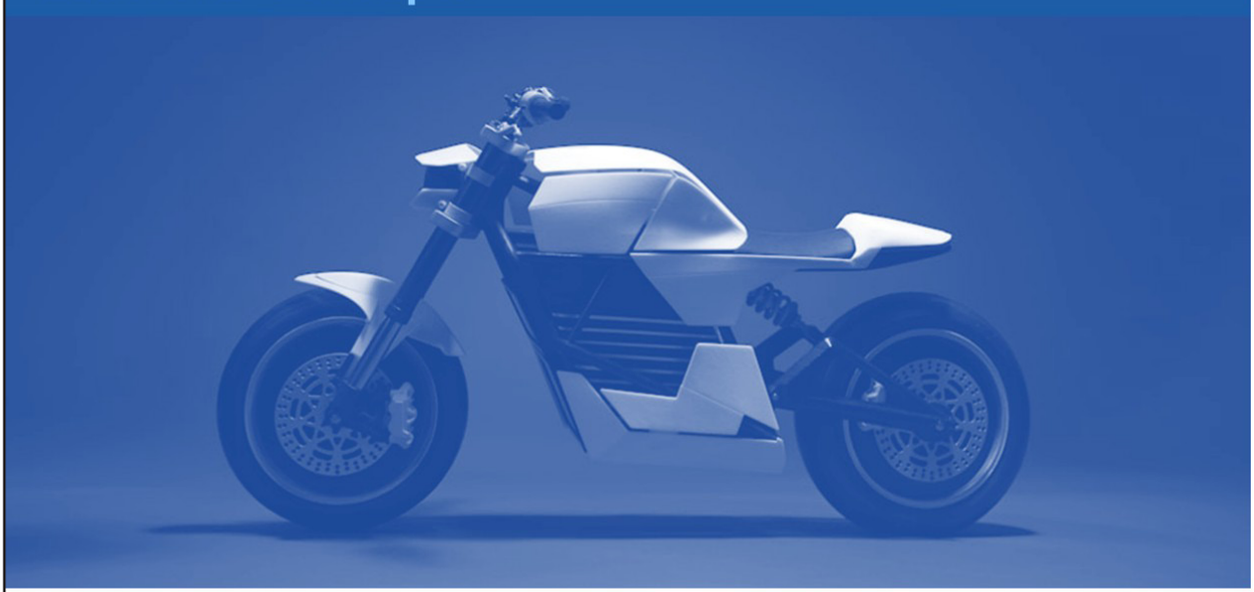
Quality

Stewardship

Controlled change

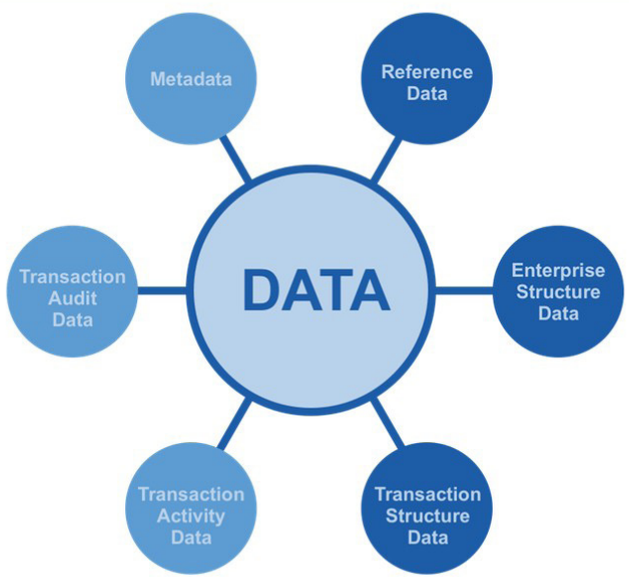
Authority

Essential Concepts



 27

Malcolm Chisholm's Taxonomy of Data



Source: Malcolm Chisholm

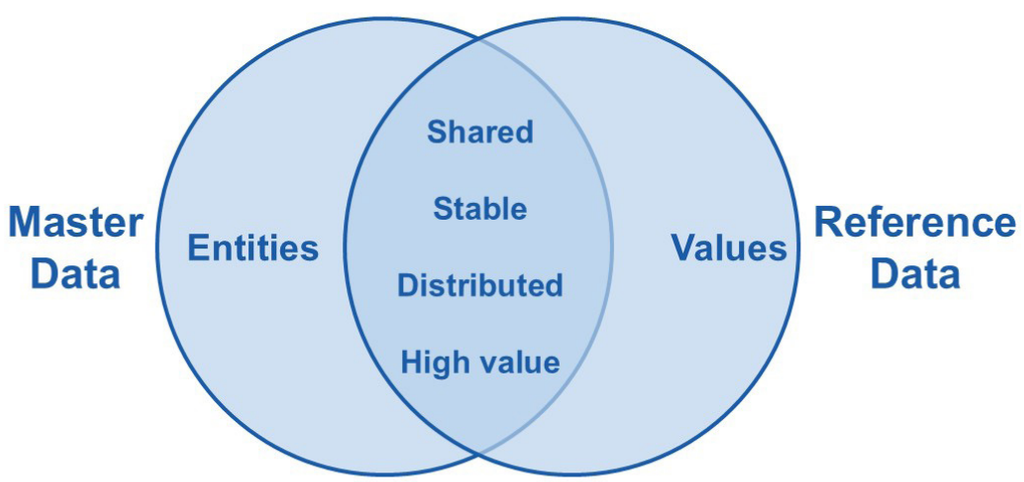
 28

Differences between Master and Reference Data

Master Data	Reference Data
Represents core business entities (Customer, Product, Supplier, Employee, Asset, Location)	Represents valid values and classifications (Country codes, Product categories, Status types)
Nouns — things the business acts upon	Adjectives / codes — how we describe or qualify data
Changes slowly, evolves over time	Changes rarely, often externally defined
Managed through MDM (Golden Records, deduplication, stewardship)	Managed through RDM (lists, hierarchies, version control)
Example: Customer "Jane Smith" with ID 12345	Example: Country Code "DE" = Germany
Answers "Who / What"	**Answers "How / Which"**

 29

Differences between Master and Reference Data



 30

Reference Data

Reference Data is
any data used to **characterize**
or **classify** other data, or to
relate data to information
external to an organization

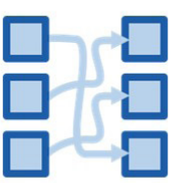
Source: Malcolm Chisholm

31

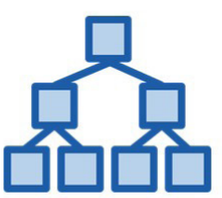
Reference Data Structure

List

Cross-Reference



Taxonomy



32

Reference Data Structure - List

alpha3-b	alpha3-t	alpha2	English name
aar		aa	Afar
abk		ab	Abkhazian
ace			Achinese
ach			Acoli
ada			Adangme
ady			Adyghe; Adygei
afa			Afro-Asiatic languages
afh			Afrihili
afr		af	Afrikaans
ain			Ainu
aka		ak	Akan
akk			Akkadian
alb	sqi	sq	Albanian
ale			Aleut
alg			Algonquian languages
alt			Southern Altai
amh		am	Amharic
ang			English, Old (ca.450-1100)
anp			Angika
apa			Apache languages
ara		ar	Arabic
arc			Official Aramaic (700-300 BCE); Imperial Aramaic (700-300 BCE)
arg		an	Aragonese
arm	hye	hy	Armenian
arn			Mapudungun; Mapuche
arp			Arapaho
art			Artificial languages
arw			Arawak
asm		as	Assamese
ast			Asturian; Bable; Leonese; Asturleonese
ath			Athapascan languages
aus			Australian languages

Source: https://datahub.io/core/language-codes

Reference Data Structure - Cross-Reference

Country	FIFA	IOC
Anguilla	AIA	
Antigua & Barbuda	ATG	ANT
Aruba	ARU	ARU
Bahamas	BAH	BAH
Barbados	BRB	BAR
Belize	BLZ	BIZ
Bermuda	BER	BER
British Virgin Islands	VGB	IVB
Canada	CAN	CAN
Cayman Islands	CAY	CAY
Costa Rica	CRC	CRC
Cuba	CUB	CUB
Curaçao	CUW	
Dominica	DMA	DMA
Dominican Republic	DOM	DOM
El Salvador	SLV	ESA
Grenada	GRN	GRN
Guatemala	GUA	GUA
Guyana	GUY	GUY
Haiti	HAI	HAI
Honduras	HON	HON

Source: https://www.rsssf.org/miscellaneous/fifa-codes.html

Reference Data Structure - Taxonomy

2022NAICSCode	LongName	TRIndustrySector	IndustrySubsector
314110	Carpet and Rug Mills	314 Textile Products	3141 Textile Furnishings Mills
314120	Curtain and Linen Mills	314 Textile Products	3141 Textile Furnishings Mills
314121	314121 Curtain and Drapery Mills	314 Textile Product	3141 Textile Furnishings Mills
314129	314129 Other Household Textile Product Mills	314 Textile Product	3141 Textile Furnishings Mills
314910	Textile Bag and Canvas Mills	314 Textile Products	3149 Other Textile Product Mills
314911	314911 Textile Bag Mills	314 Textile Product	3149 Other Textile Product Mills
314912	314912 Canvas and Related Product Mills	314 Textile Product	3149 Other Textile Product Mills
314991	314991 Rope, Cordage, and Twine Mills	314 Textile Product	3149 Other Textile Product Mills
314992	314992 Tire Cord and Tire Fabric Mills	314 Textile Product	3149 Other Textile Product Mills
314994	Rope, Cordage, Twine, Tire Cord, and Tire Fabric Mills	314 Textile Products	3149 Other Textile Product Mills
314999	All Other Miscellaneous Textile Product Mills	314 Textile Products	3149 Other Textile Product Mills
315110	Hosiery and Sock Mills	315 Apparel	3151 Apparel Knitting Mills
315111	315111 Sheer Hosiery Mills	315 Apparel	3151 Apparel Knitting Mills
315119	315119 Other Hosiery and Sock Mills	315 Apparel	3151 Apparel Knitting Mills
315120	Apparel Knitting Mills	315 Apparel	3151 Apparel Knitting Mills
315190	Other Apparel Knitting Mills	315 Apparel	3151 Apparel Knitting Mills
315191	315191 Outerwear Knitting Mills	315 Apparel	3151 Apparel Knitting Mills
315192	315192 Underwear and Nightwear Knitting Mills	315 Apparel	3151 Apparel Knitting Mills
315210	Cut and Sew Apparel Contractors	315 Apparel	3152 Cut and Sew Apparel Manufacturing
315211	315211 Men's and Boys' Cut and Sew Apparel Contractors	315 Apparel	3152 Cut and Sew Apparel Manufacturing
315212	315212 Women's, Girls', and Infants' Cut and Sew Apparel Contractors	315 Apparel	3152 Cut and Sew Apparel Manufacturing
315220	Men?s and Boys? Cut and Sew Apparel Manufacturing	315 Apparel	3152 Cut and Sew Apparel Manufacturing
315221	315221 Men's and Boys' Cut and Sew Underwear and Nightwear Manufacturing	315 Apparel	3152 Cut and Sew Apparel Manufacturing
315223	315223 Men's and Boys' Cut and Sew Shirt (except Work Shirt) Manufacturing	315 Apparel	3152 Cut and Sew Apparel Manufacturing
315224	315224 Men's and Boys' Cut and Sew Trouser, Slack, and Jean Manufacturing	315 Apparel	3152 Cut and Sew Apparel Manufacturing
315225	315225 Men's and Boys' Cut and Sew Work Clothing Manufacturing	315 Apparel	3152 Cut and Sew Apparel Manufacturing
315231	315231 Women's and Girls' Cut and Sew Lingerie, Loungewear, and Nightwear Manufacturing	315 Apparel	3152 Cut and Sew Apparel Manufacturing
315233	315233 Women's and Girls' Cut and Sew Dress Manufacturing	315 Apparel	3152 Cut and Sew Apparel Manufacturing
315234	315234 Women's and Girls' Cut and Sew Suit, Coat, Tailored Jacket, and Skirt Manufacturing	315 Apparel	3152 Cut and Sew Apparel Manufacturing
315240	Women?s, Girls?, and Infants? Cut and Sew Apparel Manufacturing	315 Apparel	3152 Cut and Sew Apparel Manufacturing
315250	Cut and Sew Apparel Manufacturing (except Contractors)	315 Apparel	3152 Cut and Sew Apparel Manufacturing
315280	Other Cut and Sew Apparel Manufacturing	315 Apparel	3152 Cut and Sew Apparel Manufacturing
315292	315292 Fur and Leather Apparel Manufacturing	315 Apparel	3152 Cut and Sew Apparel Manufacturing
315299	315299 All Other Cut and Sew Apparel Manufacturing	315 Apparel	3152 Cut and Sew Apparel Manufacturing
315990	Apparel Accessories and Other Apparel Manufacturing	315 Apparel	3159 Apparel Accessories and Other Apparel Manufacturing
315991	315991 Hat, Cap, and Millinery Manufacturing	315 Apparel	3159 Apparel Accessories and Other Apparel Manufacturing
315992	315992 Glove and Mitten Manufacturing	315 Apparel	3159 Apparel Accessories and Other Apparel Manufacturing
315999	315999 Other Apparel Accessories and Other Apparel Manufacturing	315 Apparel	3159 Apparel Accessories and Other Apparel Manufacturing
316110	Leather and Hide Tanning and Finishing	316 Leather	3161 Leather and Hide Tanning and Finishing
316210	Footwear Manufacturing	316 Leather	3162 Footwear Manufacturing
316211	316211 Rubber and Plastics Footwear Manufacturing	316 Leather	3162 Footwear Manufacturing
316212	316212 House Slipper Manufacturing	316 Leather	3162 Footwear Manufacturing
316213	316213 Men's Footwear (except Athletic) Manufacturing	316 Leather	3162 Footwear Manufacturing
316214	316214 Women's Footwear (except Athletic) Manufacturing	316 Leather	3162 Footwear Manufacturing
316219	316219 Other Footwear Manufacturing	316 Leather	3162 Footwear Manufacturing

- 314 Textile Products
  - 3141 Textile Furnishings Mills
  - 3149 Other Textile Product Mills
- 315 Apparel
  - 3151 Apparel Knitting Mills
  - 3152 Cut and Sew Apparel Manufacturing
  - 3159 Apparel Accessories and Other Apparel Manufacturing
- 316 Leather
  - 3161 Leather and Hide Tanning and Finishing
  - 3162 Footwear Manufacturing

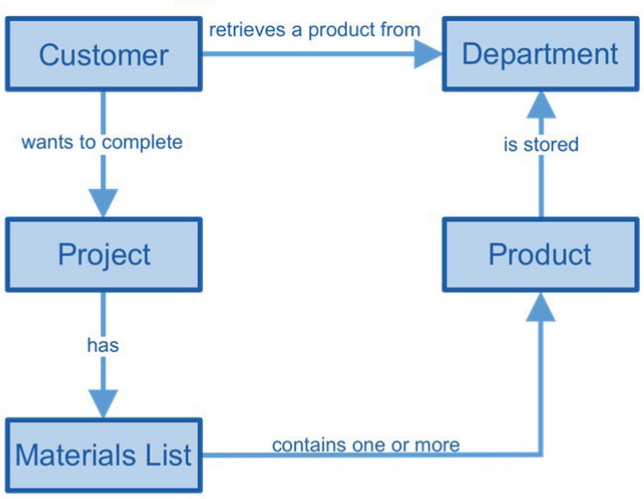
Source: https://www.epa.gov/reseller-data-dictionary-north-american-industry-classification-system-naics-data © Van Haren Publishing 35

Reference Data Structure – Taxonomy & Ontology

Taxonomy

- Tools
  - Power
  - Hand
    - Hammers
    - Wrenches
      - Adjustable
      - Pipe
      - Torque

Ontology



Source: https://www.earley.com/insights/what-difference-between-taxonomy-and-ontology-it-matter-complexity © Van Haren Publishing 36

Types of Reference Data

Proprietary

HEX product code	Product Name
00000001	XmR Control Chart
00000002	HTML Viewer
00000004	Cycle Plot
00000008	Strip Plot
00000010	Lipstick Column Chart
00000020	Lipstick Bar Chart
00000040	Lollypop Column Chart
00000080	Lollypop Bar Chart
00000100	Dumbbell Column Chart
00000200	Dumbbell Bar Chart
00000400	Merged Bar Chart
00000800	SMART KPI List
00001000	PowerGantt Chart
00002000	Milestone Trend Analysis

Industry

Chapter number	Chapter title
1	Certain infectious and/or parasitic diseases
2	Neoplasms
3	Diseases of the blood and blood-forming organs and certain disorders involving the immune mechanism
4	Endocrine, nutritional and metabolic diseases
5	Mental and behavioural disorders
6	Diseases of the nervous system
7	Diseases of the eye and adnexa
8	Diseases of the ear and mastoid process
9	Diseases of the circulatory system
10	Diseases of the respiratory system
11	Diseases of the digestive system
12	Diseases of the skin and subcutaneous tissue

Geographic

PostCode	PostcodeNum	PostcodeChar	Street	City	Municipality	Province
6651EH	6651	EH	Genechtstraat	Druten	Druten	Gelderland
1189WK	1189	WK	De Wijde Blik	Amstelveen	Amstelveen	Noord-Holland
1065VL	1065	VL	Jacob Geelstraat	Amsterdam	Amsterdam	Noord-Holland
8471RK	8471	RK	Beatrixstraat	Wolvega	Weststellingwerf	Friesland
7231JH	7231	JH	Breegraven	Warnsveld	Zutphen	Gelderland
6828AA	6828	AA	Apeldoornsestraat	Arnhem	Arnhem	Gelderland
1951HL	1951	HL	Witte Hekweg	Velsen-Noord	Velsen	Noord-Holland
1561AT	1561	AT	Noorderhoofdstraat	Krommenie	Zaanstad	Noord-Holland
9202LB	9202	LB	Donia	Drachten	Smallingerland	Friesland
4701LV	4701	LV	Burgemeester Schoonheijtstraat	Roosendaal	Roosendaal	Noord-Brabant
1741JV	1741	JV	George Breitnerstraat	Schagen	Schagen	Noord-Holland
1433BL	1433	BL	Voltastraat	Kudelstaart	Aalsmeer	Noord-Holland

Computational

Code	Currency	Spot rate €
USD	US dollar	1.1514
JPY	Japanese yen	177.57
BGN	Bulgarian lev	1.9558
CZK	Czech koruna	24.339
DKK	Danish krone	7.4676
GBP	Pound sterling	0.8765
HUF	Hungarian forint	387.35
PLN	Polish zloty	4.254
RON	Romanian leu	5.0856
SEK	Swedish krona	10.935
CHF	Swiss franc	0.9298
ISK	Icelandic krona	145.8

Reference Data Metadata

Attribute	Description
Formal name	Official, especially if external name of the Reference Data set (e.g., ISO3166-1991 Country Code List)
Internal name	Name associated with the data set within the organization (e.g., Country Codes – ISO)
Data provider	The party that provides and maintains the Reference Data set. This can be external (ISO), internal (a specific department), or external – extended (obtained from an external party but then extended and modified internally).
Data provider data set source	Description of where data provider's data sets can be obtained. This is likely a Universal Resource Identifier (URI) within or outside of the enterprise network.
Data provider latest version number	If available and maintained, this describes the latest version of the external data provider's data set where information may be added or deprecated from the version in the organization
Data provider latests version date	If available and maintained, this describes when the standard list was last updated
Internal version number	Version number of the current Reference Data set or version number of the last update that was applied against the data set
Internal version reconciliation date	Date when data set was last updated based on the external source
Internal version last update date	Date data set was last changed. This does not mean reconciliation with an external version.

Master Data

"The data that provides the **context** for business activity data in the form of **common** and **abstract concepts** that relate to the activity. It includes the details (definitions and identifiers) of **internal** and **external objects** involved in business transactions, such as customers, products, employees, vendors, and controlled domains (code values)"

Source: DAMA 2009

39

Master Data – Common examples

Involved Party

Product/Service

Account/Financial

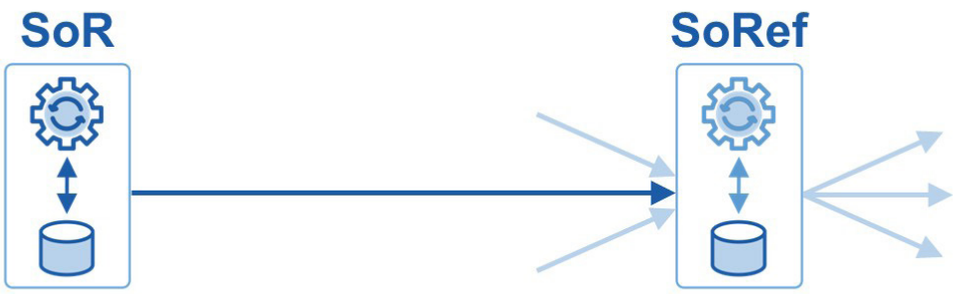
Asset

Location

Organization

40

System of Record vs. System of Reference



Authoritative source
- Created
- Maintained
- Governed

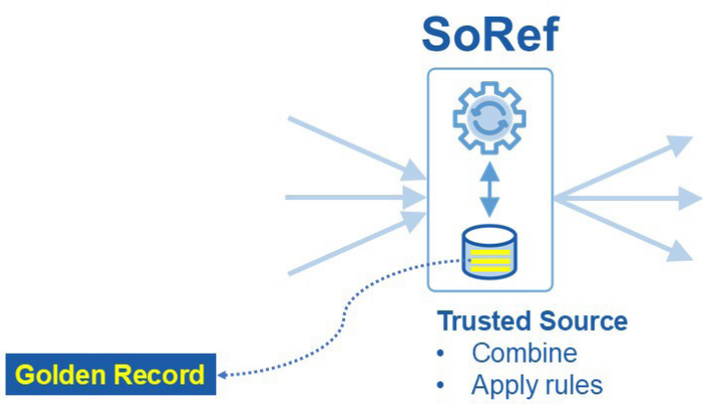
One Truth

Trusted access point
- Aggregated
- Harmonized
- Distributed

Many uses

 41

Trusted Source and Golden Record



Trusted Source
- Combine
- Apply rules
- Validate

Best currently available version of the master data

 42

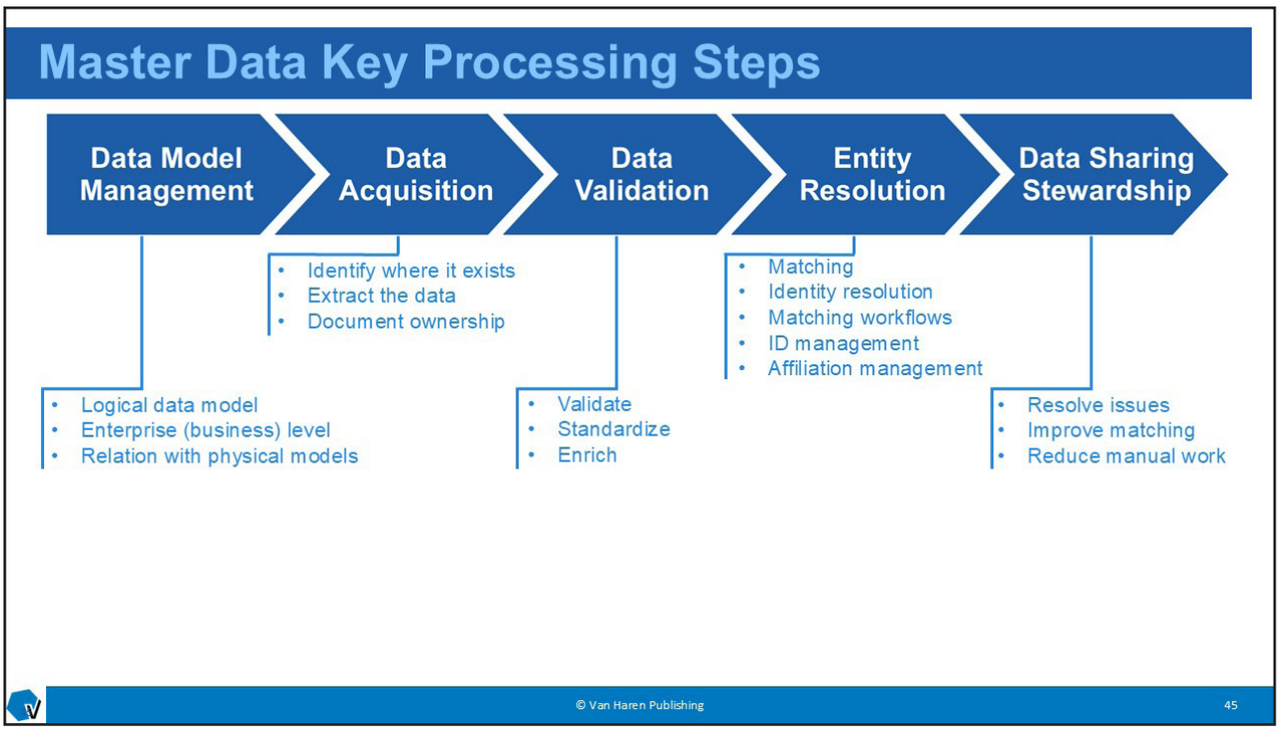
Master Data Management

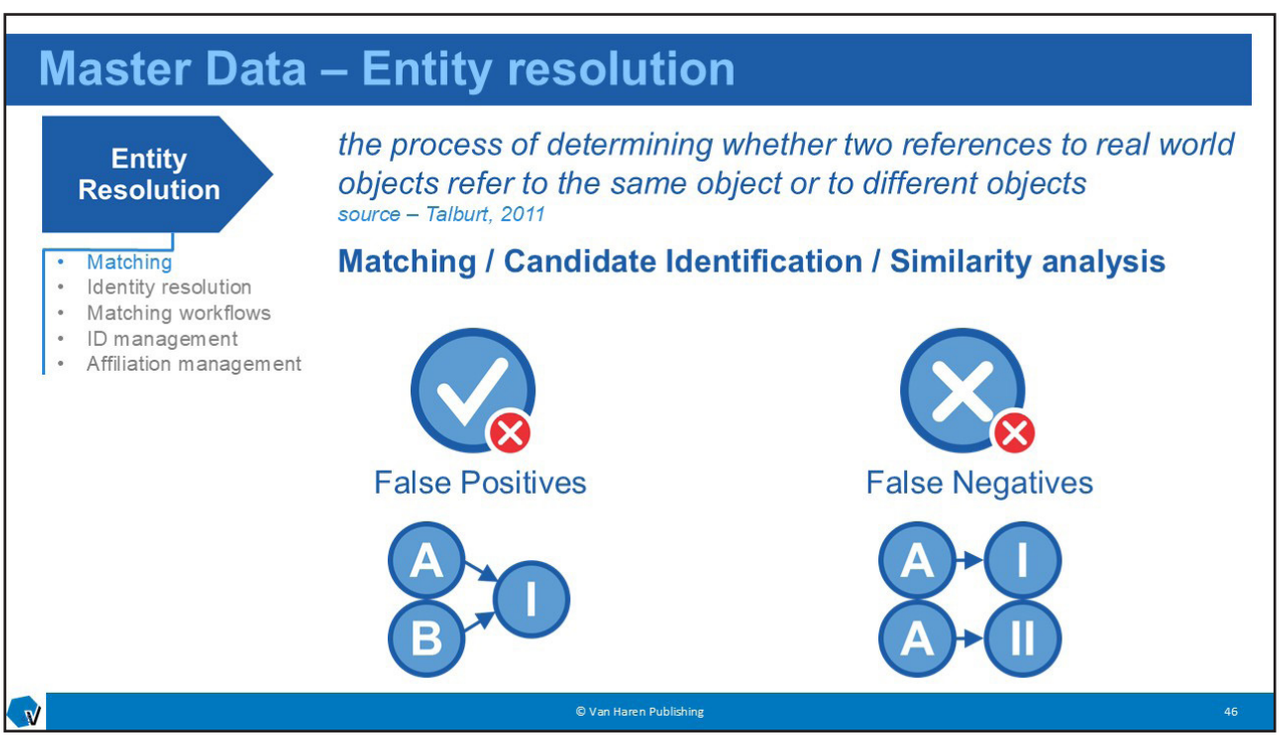
a technology-enabled discipline in which **business** and **IT** work **together** to ensure the **uniformity**, **accuracy**, **stewardship**, semantic **consistency**, and **accountability** of the enterprise's official shared Master Data assets.

Master Data is the consistent and uniform set of identifiers and extended attributes that describes the core entities of the enterprise including customers, prospects, citizens, suppliers, sites, hierarchies, and chart of accounts.

Source: https://www.gartner.com/

Master Data Lifecycle


Master Data Key Processing Steps
Data Model Management
Data Acquisition
Data Validation
Entity Resolution
Data Sharing Stewardship
• Logical data model
• Enterprise (business) level
• Relation with physical models
• Identify where it exists
• Extract the data
• Document ownership
• Validate
• Standardize
• Enrich
• Matching
• Identity resolution
• Matching workflows
• ID management
• Affiliation management
• Resolve issues
• Improve matching
• Reduce manual work
© Van Haren Publishing
45


Master Data – Entity resolution
Entity Resolution
• Matching
• Identity resolution
• Matching workflows
• ID management
• Affiliation management
the process of determining whether two references to real world objects refer to the same object or to different objects
source – Talburt, 2011
Matching / Candidate Identification / Similarity analysis
False Positives
False Negatives
A
B
I
A
I
A
II
© Van Haren Publishing
46

Master Data – Entity resolution

Entity Resolution

- Matching
- Identity resolution
- Matching workflows
- ID management
- Affiliation management

the process of determining whether two references to real world objects refer to the same object or to different objects
source – Talburt, 2011

Formalize the relationship

Source System	Source ID	Enterprise ID
CRM	CUST-00981	E12345
ERP	451-982	E12345
Marketing	00008754	E12345

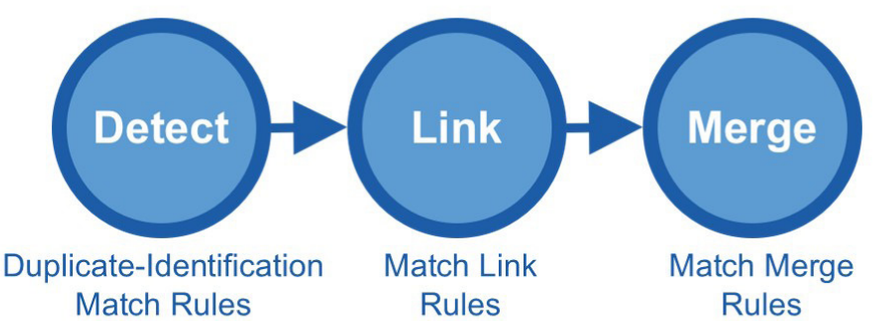
Master Data – Entity resolution

Entity Resolution

- Matching
- Identity resolution
- Matching workflows
- ID management
- Affiliation management

the process of determining whether two references to real world objects refer to the same object or to different objects
source – Talburt, 2011

Automate the process



Master Data – Entity resolution

Entity Resolution

- Matching
- Identity resolution
- Matching workflows
- ID management
- Affiliation management

the process of determining whether two references to real world objects refer to the same object or to different objects

source – Talburt, 2011

Universally known, findable, and traceable

Source System	Source ID	Enterprise ID	Status	Valid From	Valid To
CRM	CUST-00981	E12345	Active	2022-06-15	
ERP	451-982	E12345	Active	2022-07-01	
Marketing	00008754	E12345	Inactive	2021-12-31	2024-09-26

Master Data – Entity resolution

Entity Resolution

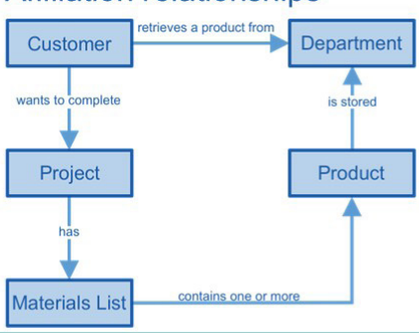
- Matching
- Identity resolution
- Matching workflows
- ID management
- Affiliation management

the process of determining whether two references to real world objects refer to the same object or to different objects

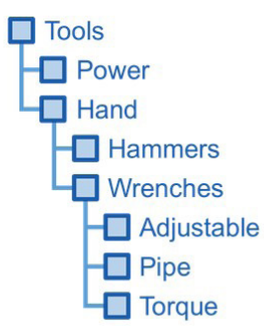
source – Talburt, 2011

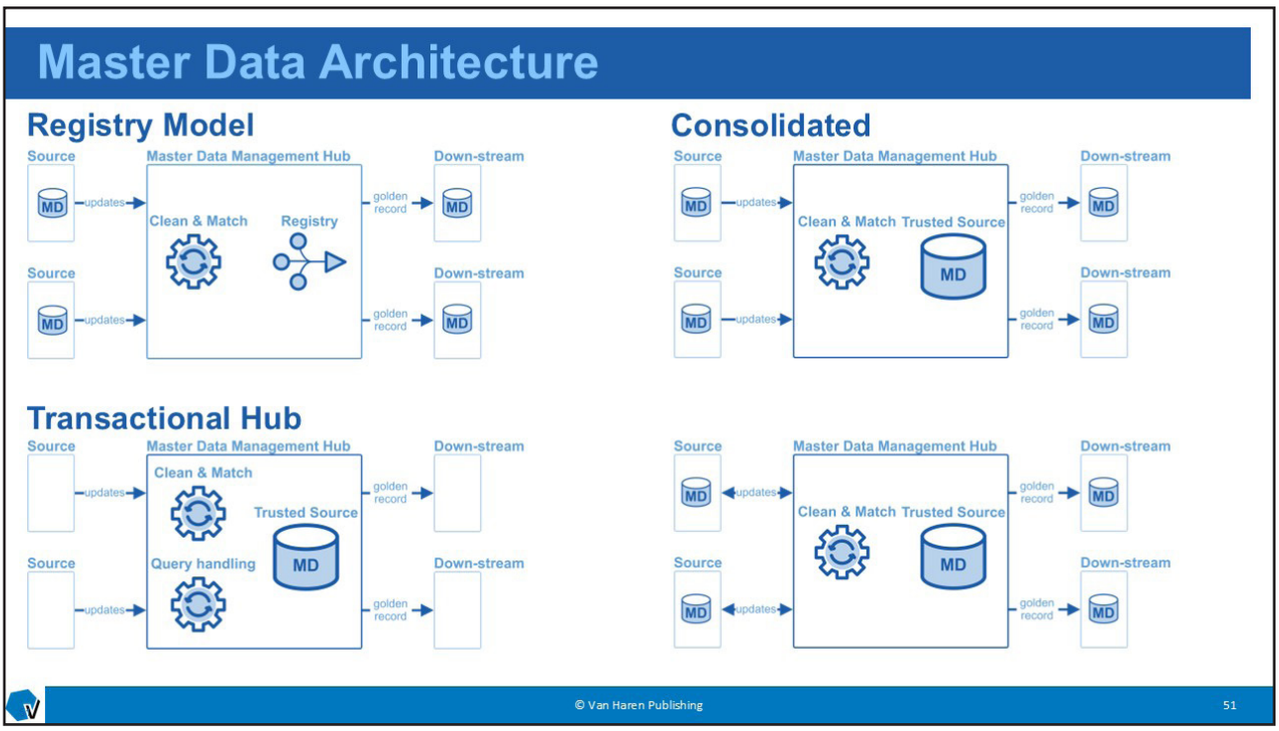
Relationships between entities

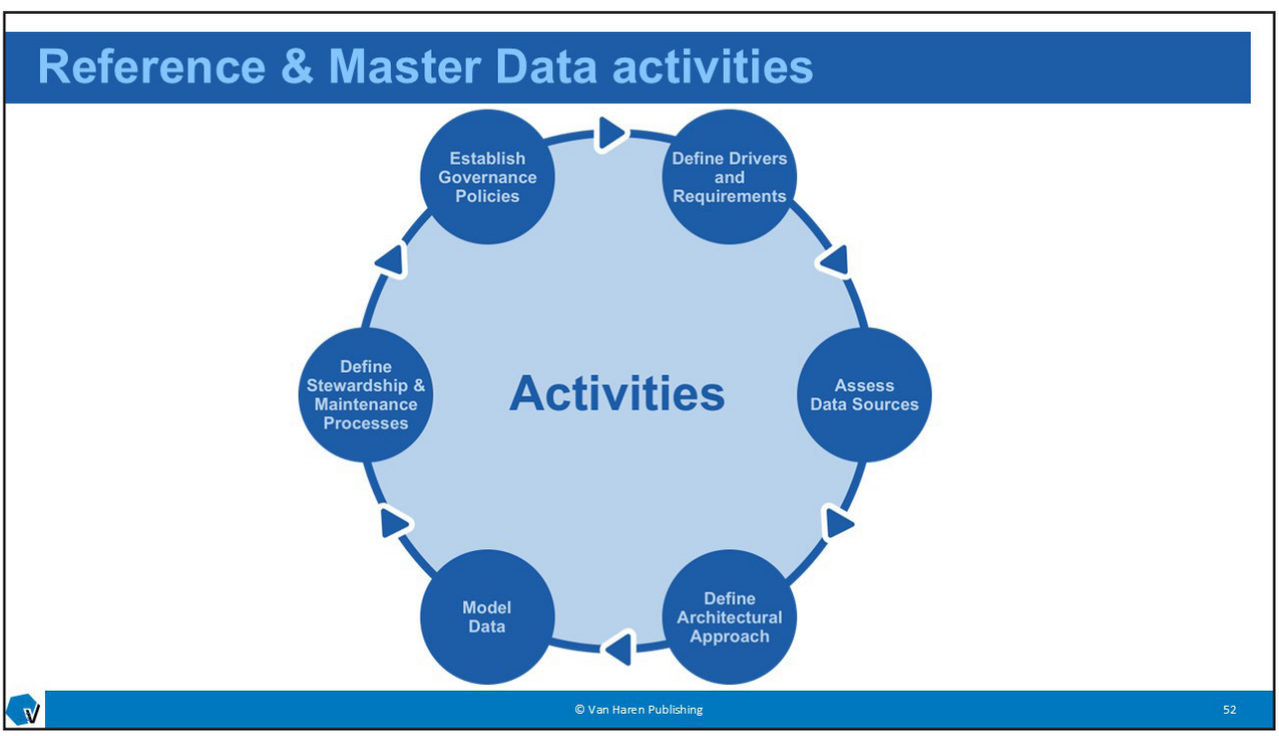
Affiliation relationships



Parent-Child relationships




Master Data Architecture
Registry Model
Source
Master Data Management Hub
Down-stream
MD
updates
Clean & Match
Registry
golden record
Consolidated
Clean & Match Trusted Source
Transactional Hub
Clean & Match
Trusted Source
Query handling
© Van Haren Publishing
51


Reference & Master Data activities
Establish Governance Policies
Define Drivers and Requirements
Assess Data Sources
Define Architectural Approach
Model Data
Define Stewardship & Maintenance Processes
Activities
© Van Haren Publishing
52

Activity - Define Drivers and Requirements

Both

- Consistency and Standardization
- Integration and Interoperability
- Data Quality and Accuracy
- Regulatory Compliance
- Analytics and Insight
- Cost and Efficiency

Reference Data

- Standardization
- Regulatory Alignment
- Change Management
- Semantic Consistency
- Version Control
- Accessibility

Master Data

- Single View / "Golden Record"
- Operational Efficiency
- Customer / Supplier 360° View
- Data Quality & Matching
- Governance of Shared Entities
- Integration with Transactions

 53

Activity – Assess Data Sources

Both

- Inventory of Sources
- Profiling & Quality Assessment
- Lineage & Ownership Analysis
- Redundancy & Overlap Detection
- Authority Evaluation
- Risk & Compliance Review

Reference Data

- Identify all internal and external lists
- Compare to industry or regulatory standards.
- Check for semantic drift
- Evaluate update mechanisms
- Determine ownership and stewardship

Master Data

- Locate all applications where key entities are created/ updated/ used.
- Evaluate data quality metrics
- Examine identifier structures
- Document integration touchpoints
- Identify data conflicts

 54