

Data Concept System for Data Quality Dimensions (DCS)

Research Paper



Colophon

Authors:

Andrew Black (Van Nederpelt & Black)

Peter van Nederpelt (Van Nederpelt & Black)

Version: See Version History

Printed on 5-9-2020 11:11

© DAMA NL Foundation. All rights reserved.

Table of contents

Tables	9
Figures	9
About the authors	10
1. Introduction	11
1.1 Importance	11
1.2 Scope	11
1.3 Research question	11
1.4 Target group	11
1.5 Background, ownership and management	11
1.6 Release policy	11
1.7 Reading guide	12
2. Methodology	13
2.1 Step 1: Collect input	13
2.2 Step 3: Evaluate data concepts in current definitions of data quality dimensions	13
2.3 Step 4: Evaluate current data concepts and data concept definitions	13
2.4 Step 5: Compose a data concept system	13
3. Results	14
3.1 Result step 1: Collect input	14
3.2 Result step 2: Evaluate data concepts in current definitions of data quality dimensions	14
3.3 Result step 3: Evaluate current data concepts and data concept definitions	15
3.4 Result step 4: Compose a data concept system	18
4. Conclusions and discussion	22
Appendix 1: Sources	23
5.1 Sources of definitions of data quality dimension	23
5.2 Sources of definitions of data concepts	23
Appendix 2: Frequencies	25
Appendix 3: Criteria for data concepts and data concept definitions	27
5.3 Criteria for data concepts	27
5.4 Criteria for data concept definitions	27
Appendix 4: Concepts and their definitions	28
1. Attribute	29
1.1 Definitions	29
1.2 Analogies and frequencies	29

1.3	Relationships	30
1.4	Notes	30
2.	Characteristic	31
2.1	Definitions	31
2.2	Analogies	31
2.3	Relationships	31
2.4	Notes	31
3.	Composition of a dataset	32
3.1	Definitions	32
3.2	Analogies and frequencies	32
3.3	Relationships	32
3.4	Notes	32
4.	Concept	33
4.1	Definitions	33
4.2	Analogies	33
4.3	Relationships	33
4.4	Notes	33
5.	Concept system	34
5.1	Definitions	34
5.2	Analogies and frequencies	34
5.3	Relationships	34
5.4	Notes	34
6.	Data	35
6.1	Definitions	35
6.2	Analogies	35
6.3	Relationships	35
6.4	Notes	35
7.	Data category	36
7.1	Definitions	36
7.2	Analogies and frequencies	36
7.3	Notes	36
8.	Data concept	37
8.1	Definitions	37
8.2	Analogies	37
8.3	Relationships	37
8.4	Notes	37

9. Data element	38
9.1 Definitions	38
9.2 Analogies	38
9.3 Relationships	38
9.4 Note	39
10. Data element specification	40
10.1 Definitions	40
10.2 Analogies and frequencies	40
10.3 Relationships with other data concepts	40
10.4 Notes	40
11. Data item	41
11.1 Definitions	41
11.2 Analogies and frequencies	41
11.3 Relationships with other data concepts	41
11.4 Notes	41
12. Data pattern	42
12.1 Definitions	42
12.2 Analogies and frequencies	42
12.3 Notes	42
13. Data lineage	43
13.1 Definitions	43
13.2 Analogies and frequencies	43
13.3 Relationships with other data concepts	43
13.4 Notes	43
14. Data file	44
14.1 Definitions	44
14.2 Analogies and frequencies	44
14.3 Relationships with other data concepts	44
14.4 Notes	44
15. Data value	45
15.1 Definitions	45
15.2 Analogies and frequencies	45
15.3 Relationships with other data concepts	45
15.4 Notes	45
16. Data value, initial	46
16.1 Definitions	46

16.2 Analogies and frequencies	46
16.3 Relationships with other data concepts	46
16.4 Notes	46
17. Database	47
17.1 Definitions	47
17.2 Analogies	47
17.3 Relationships with other data concepts	47
17.4 Note	47
18. Dataset	48
18.1 Definitions	48
18.2 Analogies	48
18.3 Relationships with other data concepts	48
18.4 Note	48
19. Dataset availability	49
19.1 Definitions	49
19.2 Analogies	49
19.3 Relationships with other data concepts	49
19.4 Note	49
20. Dataset composition	50
20.1 Definitions	50
20.2 Analogies	50
20.3 Relationships with other data concepts	50
20.4 Note	50
21. Definition	51
21.1 Definitions	51
21.2 Analogies and frequencies	51
21.3 Relationships with other data concepts	51
21.4 Notes	51
22. Dimension	52
22.1 Definitions	52
22.2 Analogies and frequencies	52
22.3 Relationships with other data concepts	52
22.4 Notes	52
23. Entity type	53
23.1 Definitions	53
23.2 Analogies and frequencies	53

23.3	Relationships with other data concepts	53
23.4	Notes	53
24.	Format	54
24.1	Definitions	54
24.2	Analogies and frequencies	54
24.3	Relationships with other data concepts	54
24.4	Notes	54
25.	Master data	55
25.1	Definitions	55
25.2	Analogies	55
25.3	Relationships with other concepts	55
25.4	Notes	55
26.	Metadata	56
26.1	Definitions	56
26.2	Analogies and frequencies	56
26.3	Relationships with other data concepts	56
26.4	Notes	56
27.	Object	57
27.1	Definitions	57
27.2	Analogies and frequencies	57
27.3	Relationships with other data concepts	57
27.4	Notes	57
28.	Object type	58
28.1	Definitions	58
28.2	Analogies and frequencies	58
28.3	Relationships with other data concepts	58
28.4	Notes	58
29.	Property	59
29.1	Definitions	59
29.2	Analogies and frequencies	59
29.3	Relationships with other data concepts	59
29.4	Notes	59
30.	Transactional data	60
30.1	Definitions	60
30.2	Analogies and frequencies	60
30.3	Relationships with other data concepts	60

30.4	Notes	60
31.	Real value	61
31.1	Definitions	61
31.2	Analogies and frequencies	61
31.3	Relationships with other data concepts	61
31.4	Notes	61
32.	Record	62
32.1	Definitions	62
32.2	Analogies and frequencies	62
32.3	Relationships with other data concepts	62
32.4	Notes	62
33.	Reference data	63
33.1	Definitions	63
33.2	Analogies and frequencies	63
33.3	Relationships with other data concepts	63
33.4	Notes	63
34.	Statistical output	64
34.1	Definitions	64
34.2	Analogies and frequencies	64
34.3	Relationships with other data concepts	64
34.4	Notes	64
35.	Value domain	65
35.1	Definitions	65
35.2	Analogies and frequencies	65
35.3	Relationships with other data concepts	65
35.4	Notes	65
	Version history	66

Tables

Table 1: Definitions of data quality dimensions (frequencies)

Table 2: Definitions of data concepts

Table 3: Frequencies of data concepts in definitions of dimensions of data quality

Table 4: Criteria for data concepts

Table 5: Criteria for definitions of data concepts

Figures

Figure 1: Data concept system DDQ

Figure 2: Data concepts in a data model

Figure 3: Relationship between data concepts and dimensions

Figure 4: Artists impression of the real world and data world

About the authors

Andrew Black

Andrew Black's early career was as an accountant in financial audit where he specialised in the audit of automated accounting systems. At Philips Corporate IT he joined a group working on standardisation of metadata in the form of detailed specifications of data element types in a corporate data dictionary. Here he was inspired by the way such standardised metadata provides a solid foundation for data quality. He went on to apply the rather theoretical corporate experience to more practical data management at Philips Semiconductors and NXP where he had experience of product data management, data migration, data cleansing and a data dictionary to enable integration of customer facing product data.

Peter van Nederpelt

Peter van Nederpelt was 17 years employed at Statistics Netherlands as quality officer and operational auditor. He dedicated his work for a great deal to quality management and published a report about quality dimensions of statistical output. Peter is author of the book Object Oriented Quality and Risk Management, which introduces a new way of looking at the data concept of quality dimensions. He was several years active in the Working Group of Quality of the European Commission (Eurostat). During his last years at Statistics Netherlands, he was involved in the implementation of ISO 9001 and performed as lead auditor internal audits as required by this standard. He was author of DAMA's Code of Information Quality: an audit framework.

1. Introduction

1.1 Importance

Data Quality is an important subject in data management. It may be seen as the primary goal of data management. Data quality is measured in terms of dimensions. In the literature many dimensions of data quality have been distinguished. These dimensions refer both to real world concepts and to concepts in the data world, such as data values, records and datasets.

Concepts and their relationships in a particular field of knowledge form a concept system. Concepts and concept systems are a way to model knowledge.

To formulate a consistent set of definitions of data quality dimensions, it is necessary to choose a suitable data concept system as basis for these definitions.

1.2 Scope

The scope of this report is a data concept system that is relevant in the context of data quality dimensions.

1.3 Research question

Our research question is:

What is a suitable data concept system for definitions of data quality dimensions?

Sub questions are:

- What data concepts are used in current definitions of data quality dimensions and how frequently are they used?
- What are criteria for selection of data concepts and their definitions for the data concept system?
- Which data concepts meet the criteria?
- Which definitions meet the criteria?
- What are the relationships between these data concepts?

1.4 Target group

The report is meant for everyone who is interested in data quality.

1.5 Background, ownership and management

The reason for compiling this report is that the authors wanted to make the Code for Information Quality consistent with DMBOK2 (2017). This applies in particular to the definitions of the dimensions of data quality.

It turned out that the definitions in DMBOK2 had not yet been elevated to the status of a standard and were sometimes open to improvement. This prompted the authors to set up the DAMA_NL Data Quality working group to improve and harmonize these definitions.

Data quality dimensions should be based on solid ground. This prompted the authors to set up a report on a data concept system for data quality dimensions.

1.6 Release policy

The first final version of the report will be published in 2020. New versions will be compiled as needed. Proposals for changes can be made via info@dama-nl.org or the authors info@vannederpeltblack.nl.

1.7 Reading guide

Chapter 2 explains the steps taken to compose a data concept system for data quality dimensions.

Chapter 3 demonstrates the results of the research.

Chapter 4 contains the conclusion on whether the research questions are answered.

Appendix 1 lists the sources used in the research.

Appendix 2 demonstrates the frequency with which data concepts are used in current definitions of data quality dimensions.

In Appendix 3 criteria for data concepts and definitions of data concepts are drawn up.

Appendix 4 is a compilation of data concepts and their definitions. In this appendix a selection of data concepts and definitions is made.

2. Methodology

This chapter describes the input used and the steps taken.

2.1 Step 1: Collect input

For this research, the following input has been used:

- Sources that provide definitions of dimensions of data quality.
- Sources that provide definitions of data concepts in the field of data.
- A standard of requirements for definitions and data concept systems.

2.2 Step 3: Evaluate data concepts in current definitions of data quality dimensions

- Collect definitions of data quality dimensions from well-known sources.
- Determine which data concepts are used in these definitions.
- Count the frequencies of the data concepts.

2.3 Step 4: Evaluate current data concepts and data concept definitions

- Define criteria for data concepts
- Define criteria for data concept definitions.
- Collect data concepts and data concept definitions that are relevant for use in definitions of data quality dimensions.
- Select data concepts that meet the criteria and consider the frequency that the data concept occurs in current definitions of data quality dimensions.
- Select data concept definitions that meet the criteria or reformulate them.

2.4 Step 5: Compose a data concept system

- Determine the relationships between the data concepts.
- Draw a diagram of the data concept system.

3. Results

3.1 Result step 1: Collect input

The result of step 1 is a list of sources that provide:

- Definitions of dimensions of data quality.
- Definitions of data concepts related to data.
- A standard of requirements for definitions and data concept systems.

See appendix 1.

3.2 Result step 2: Evaluate data concepts in current definitions of data quality dimensions

Ten respected sources of definitions of data quality dimensions are consulted. These sources contain 138 definitions of data quality dimensions, which contain **57** different data concepts. See appendix 2.

Table 1: Definitions of data quality dimensions (frequencies)

Number of sources	10
Number of definitions	138
Number of different data concepts	57

3.3 Result step 3: Evaluate current data concepts and data concept definitions

Step 3 resulted in a set of selected data concepts and definitions of these data concepts. Each word that appears in **bold** in the definition of a data concept is a data concept defined elsewhere in table 2. Relationships with other data concept not mentioned in the definition are added. This way the coherence between the data concepts are made visible.

A distinction is made between data concepts in de real world (purple) and data world (yellow). Other concepts (white) are more general.

Table 2: Data concept system

Data concept	Definition	Source	Relationships with other concepts
Attribute	A characteristic of an entity type about which the organisation wishes to hold information.	-	Distinguishes entity type Is specified by its name, definition, classification and format .
Characteristic	Distinguishing feature	ISO 9000	-
Composition of a dataset	The way in which a dataset is made up.	-	
Concept	Unit of knowledge created by a unique combination of characteristics	ISO 1087	-
Concept system	A set of concepts structured according to the relations among them.	ISO 704	-
Data	A representation of facts, concepts, or instructions in a formalized manner, suitable for communication, interpretation, or processing by humans or by automatic means. (ISO 2382-4).	In: ISO 11179	-
Data category	A classification of data according to the purpose for which it is used.	-	
Data concept	A concept related to data	-	Has associated dimensions
Data file	Data stored on a computer as one unit with one name.	Cambridge 2020	Is part of a dataset .
Data item	One occurrence of an attribute	-	Contains data value
Data lineage	Metadata that identifies the sources of data and the transformations through which it has passed up to the point of consumption.	-	-
Data pattern	A series of data that repeats in a recognizable way.	Investopedia	-

Data concept	Definition	Source	Relationships with other concepts
Data value	The value of a data item .	-	Is contained in data item Forms part of record Is within value domain Represents a property of an object
Dataset	Any organized collection of data .	Early 2011	Is composed of data files
Dataset availability	The degree to which a dataset can be consulted or retrieved by data consumers or processes.	-	Is a characteristic of a dataset .
Dataset composition	The way in which a dataset is made up.	-	-
Definition	Representation of a concept by an expression that describes it and differentiates it from related concepts	ISO 1087	-
Dimension	Measurable characteristic .	DAMA 2017	Is associated with a data concept .
Entity type	A thing of significance about which the organisation wishes to hold information	Hay 2013	Is distinguished by attributes Describes object
Initial data value	A provisional data value that will be updated by a more accurate value.	-	Is a specification of a data value .
Format	A combination of datatype, unit of measure and character set.	-	Is part of the specification of an attribute .
Metadata	Data that defines and describes other data .	ISO 11179	Is a an instance of data category .
Master Data	Data held by an organization which describe object types that it needs to reference in order to perform its transactions.	-	Is an instance of data category .
Object	Anything perceivable or conceivable.	ISO 9000	Is described by entity type Is characterised by properties Is represented by records
Property	A feature of an object .	ISO 1087	Characterises object Is recorded by data value Actually, has real value

Data concept	Definition	Source	Relationships with other concepts
Register	A dataset designated by the government in which vital data about citizens, residents, companies, institutions, vehicles, topography, buildings, and addresses can be centrally maintained.	-	Is an instance of data category .
Statistical output	Output from a statistical process.	-	Is an instance of data category .
Transactional data	Data that describes an event that takes place as an organization conducts its business.	-	Is an instance of data category .
Real value	The real-life value of a property of an object .	-	Expresses an instance of a property .
Reference data	Data used to categorize other data.	-	Is an instance of data category .
Record	A logically related set of data values that represent a (real-world) object	-	Forms part of data file Is composed of data values
Value domain	A set of permissible values of an attribute .	-	Includes data value

3.4 Result step 4: Compose a data concept system

The data concept system relevant for dimensions of data quality (DDQ) is shown in figure 1.

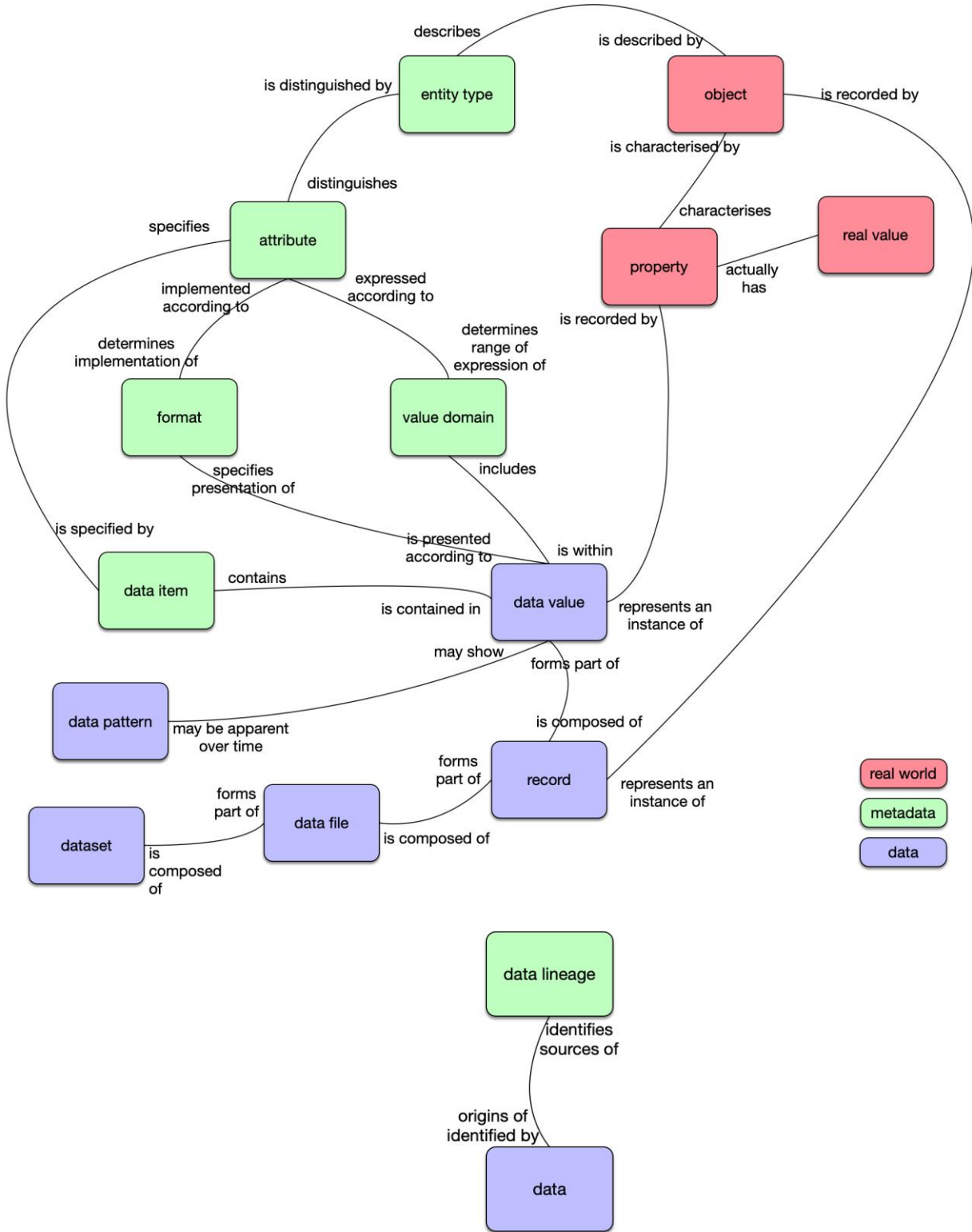


Figure 1: Data concept system DDQ

A set of data concepts is shown in a data model in figure 2.

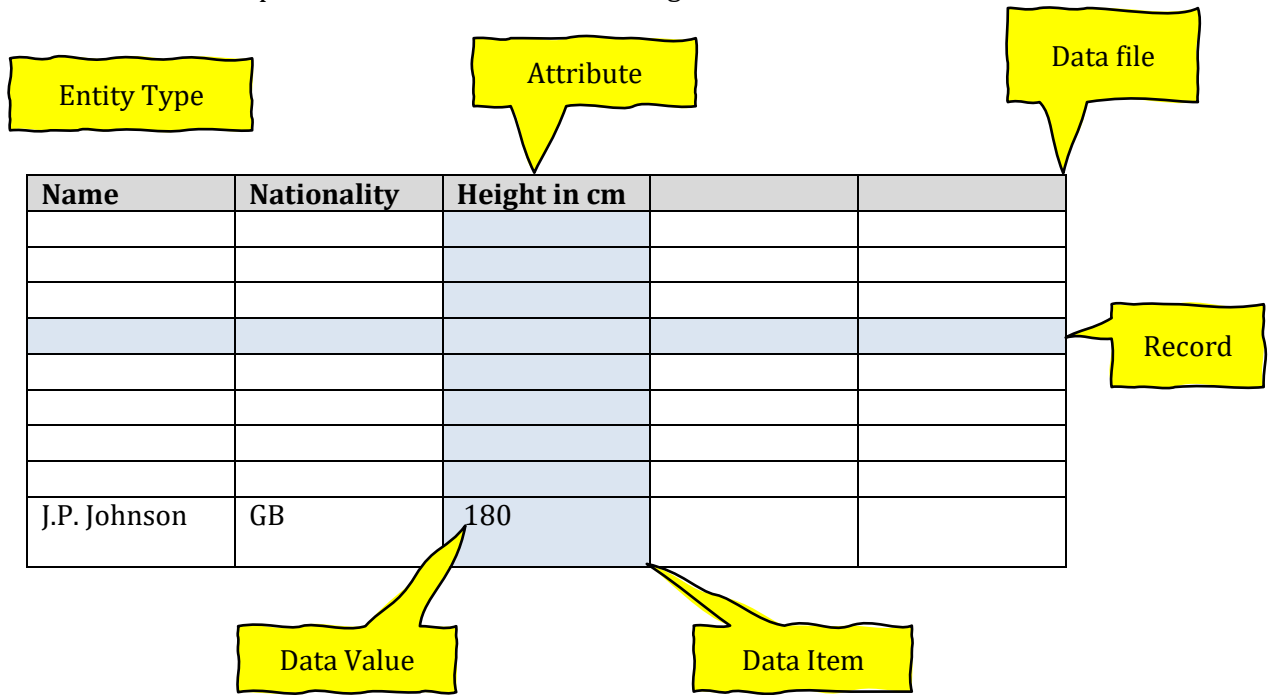


Figure 2: Data concepts in a data model

Figure 3 shows that a dimension is associated with a data concept. The definition of a dimension of data quality describes the combination of a dimension and a data concept.

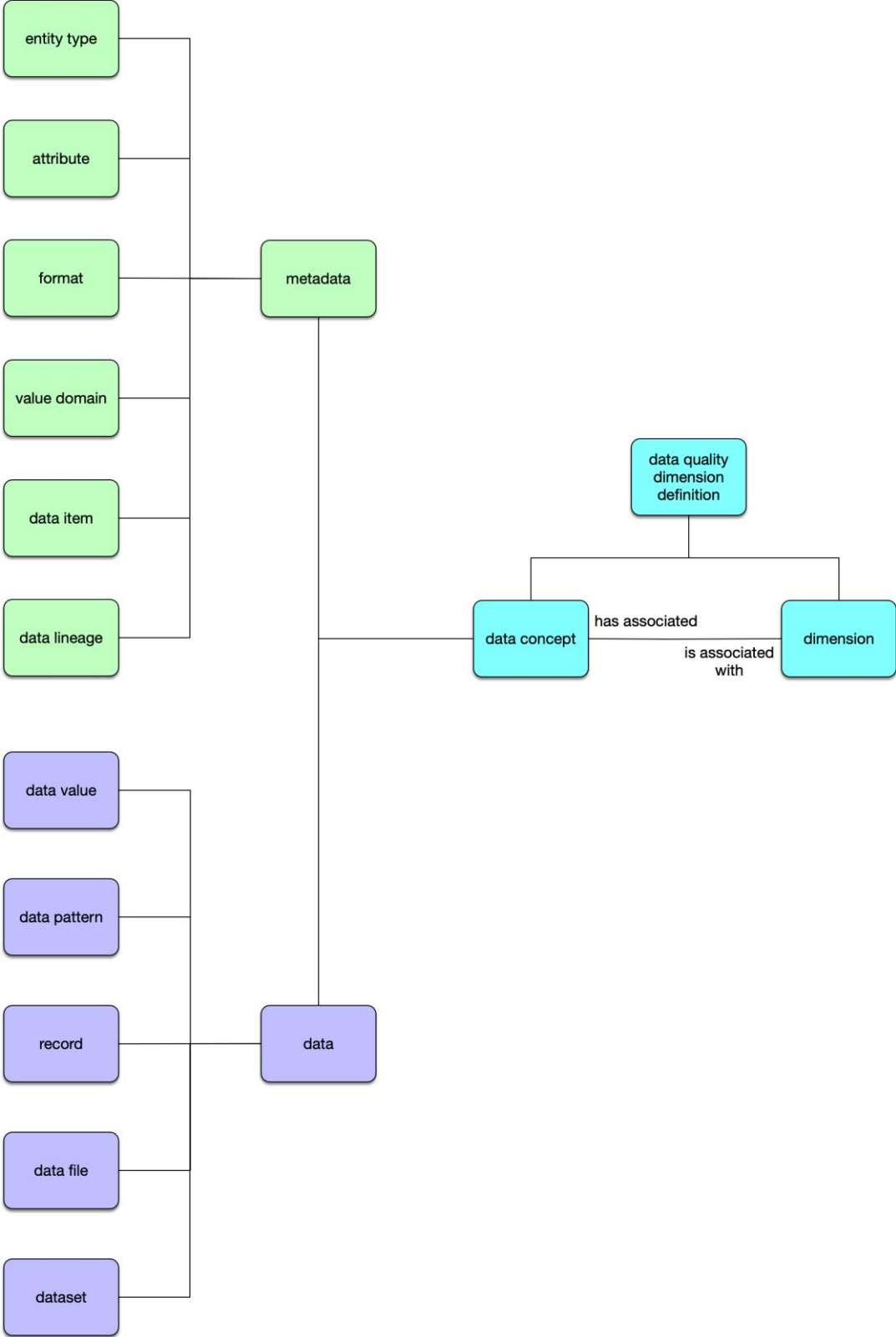


Figure 3: Relationship between data concepts and dimensions

Figure 4 is an artist's impression of the real world and data world.

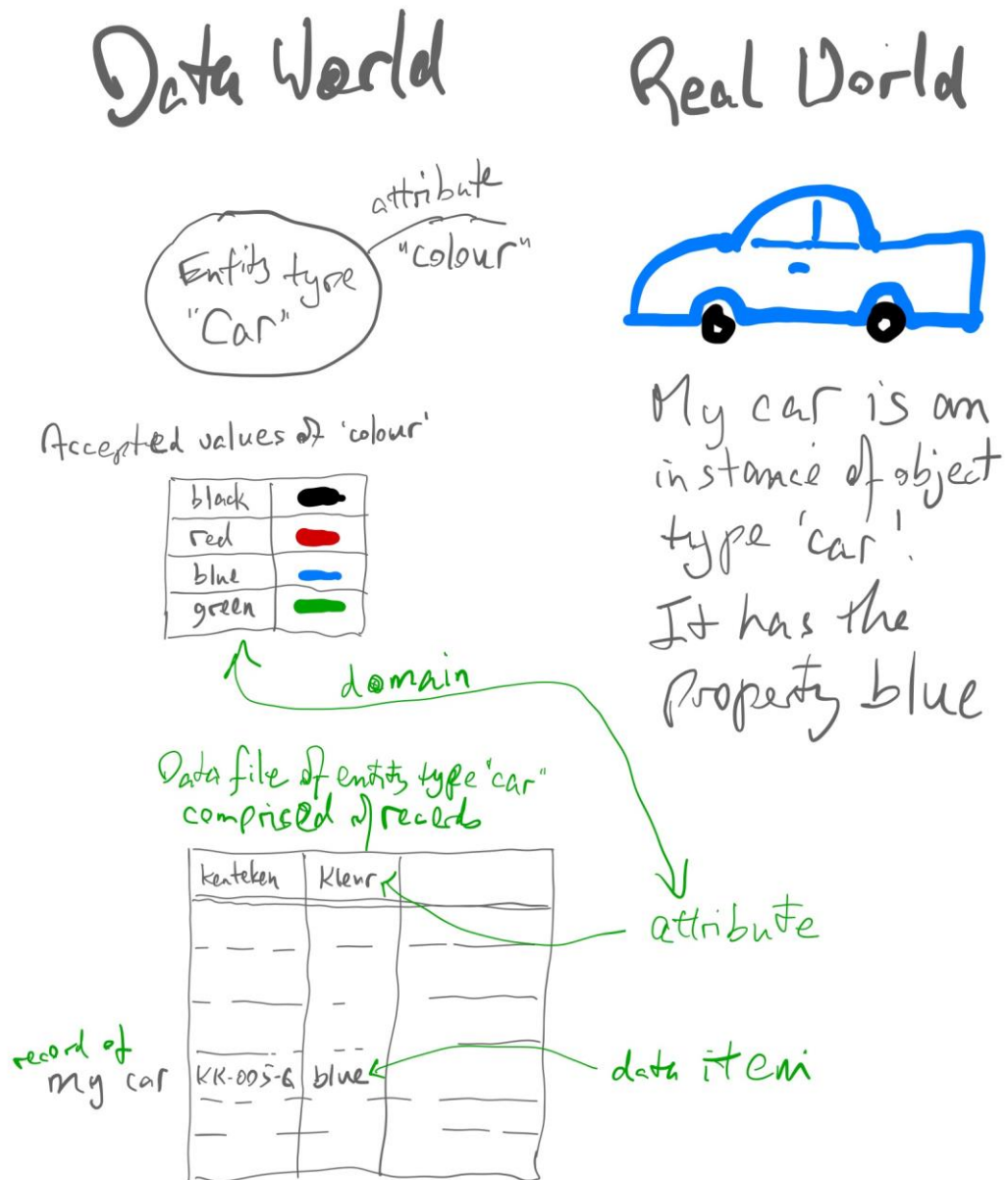


Figure 5: Artists impression of the real world and data world

4. Conclusions and discussion

Our research question was “What is a suitable data concept system for definitions of data quality dimensions?”

The answer to this question is shown in section 3.3, where the data concepts are defined. In section 3.4 the data concept system is illustrated by diagrams.

The result will be used as input for the report Dimensions of Data Quality (DDQ).

Appendix 1: Sources

5.1 Sources of definitions of data quality dimension

- Brackett, M. H. (2012). *Data Resource Integration: Understanding and Resolving a Disparate Data Resource* (1st ed.). Bradley Beach, NJ: Technics Publications, LLC.
- CDDQ. (n.d.). List of Conformed Dimensions of Data Quality. Retrieved from <https://dimensionsofdataquality.com/alldimensions>
- Daas, P.J.H., Ossen, S.J.L., & Tennekes, M. (2010). *Determination of Administrative Data Quality: Recent results and new developments*. Retrieved from Q2010 website: https://q2010.stat.fi/media/presentations/special-session-34/daas_ossen_tennekes_q2010_paper_session34_piet_daas_paper.pdf
- DAMA (2017). DAMA-DMBOK. *Data Management Body of Knowledge*. 2nd Edition. Technics Publications Llc. August 2017.
- DAMA-UK (2013). *The six primary dimensions for data quality assessment*. October 2013.
- Earley, S. (2011). *The DAMA Dictionary of Data Management* (2nded.). NJ: Technics Publications Llc.
- English, L. P. (1999). *Improving Data Warehouse and Business Information Quality: Methods for Reducing Costs and Increasing Profits*. Hoboken, NJ: Wiley.
- Eurostat. (2015). *ESS Handbook for Quality Reports*. (2015). Brussels, Belgium: Eurostat.
- Everest. (2010).
- Fisher, Craig; Lauria, Eitel; Chengalur-Smith, Shobha; Wang, Richard (2011). *Introduction to Information Quality*. Bloomington: AuthorHouse.
- ISO 25012. (n.d.). Retrieved from <https://iso25000.com/index.php/en/iso-25000-standards/iso-25012>
- Jayawardene, J., Sadiq, S., & Indulska, M. (2015). An analysis of data quality dimensions. *Computer Science*. Retrieved from <https://pdfs.semanticscholar.org/3d9a/c49f03b3e4bebae7c0e7eb20d7fde7222a9c.pdf>
- Nederpelt, P.W.M. van (2009). *Checklist Quality of Statistical Output*. Den Haag/Heerlen: Centraal Bureau voor de Statistiek.
- Redman, T. C. (1996). *Data Quality for the Information Age*. Artech House on Demand.
- Wang, R.Y. and Strong, D. (1996). *Beyond Accuracy: What data quality means to Data Consumers*. *Journal of Management Information Systems*, 1996. 12(4): p. 5 – 34.

5.2 Sources of definitions of data concepts

- Brackett, M.H. (2012). *Data Resource Integration: Understanding and Resolving a Disparate Data Resource* (1st ed.). Bradley Beach, NJ: Technics Publications, LLC.
- Business Directory. (2020, 1 mei). Geraadpleegd op 2 mei 2020, van <http://www.businessdictionary.com/definition/conceptual-framework.html>
- Cambridge English dictionary. (2020, August 1). *Meanings & definitions*. Cambridge Dictionary | English Dictionary, Translations & Thesaurus. <https://dictionary.cambridge.org/dictionary/english>
- DAMA (2017). DAMA-DMBOK. *Data Management Body of Knowledge*. 2nd Edition. Technics Publications Llc. August 2017.
- Earley, S. (2011). *The DAMA Dictionary of Data Management* (2nd ed.). NJ: Technics Publications Llc.
- ECCMA (2020). *EOTD search dictionary*. https://eccma.org/private/advanced_search_dictionary.php?
- Hay, D. (2013). *Data model patterns: Conventions of thought*. Addison-Wesley.

- Humbley, J., Budin, G., & Laurén, C. (2018). *Languages for Special Purposes*. Berlin, Germany: De Gruyter.
- Investopedia. (2015, January 7). *Patterns vs. trends: What's the difference?* <https://www.investopedia.com/ask/answers/010715/what-are-differences-between-patterns-and-trends.asp>
- ISO 1087 (2019). *Terminology work and terminology science — Vocabulary*. Vernier, Switzerland: ISO.
- ISO 11179-1 (1999). *Information technology — Specification and standardization of data elements — Part 1: Framework for the specification and standardization of data elements*. ISO
- ISO 21961 (2003). *Space data and information transfer systems – Data entity dictionary specification language (DEDSL) – Abstract syntax*. Retrieved from <https://www.iso.org/obp/ui#iso:std:iso:21961:ed-1:v1:en>.
- ISO 704 (2009). Technical Committee ISO/TC 37; Terminology and other language and content resources. Subcommittee SC 1; Principles and Methods. (2009). *Terminology Work – Principles and Methods*. Vernier, Switzerland: ISO.
- ISO 9000:2015 (2015). *Quality management systems – Fundamentals and vocabulary*. Delft: NNI.
- ISO 9001:2015 (2015). *Quality management systems – Requirements*. Delft: NNI.
- Jonker R. (2020, May 6). Terminologie. Retrieved from <https://labyrinth.rienkjonker.nl/lexicon/terminologie>
- Lexico Dictionaries (2020, August 1). *Definitions, meanings, synonyms, and grammar by Oxford dictionary on Lexico.com*. Lexico Dictionaries | English. <https://www.lexico.com>
- McGilvray, D. (2008). *Executing Data Quality Projects: Ten Steps to Quality Data and Trusted Information*. Morgan Kaufman Publishers.
- Merriam-Webster. (2020, May 7). Dictionary by Merriam-Webster: America's most-trusted online dictionary. Retrieved from <https://www.merriam-webster.com/dictionary>
- Nederpelt, P.W.M. van (2012). *Object-oriented Quality and Risk Management. A practical method for quality and risk management*. New York/Alphen den Rijn: Lulu/MicroData.
- Oxford. (2020, May 7). Oxford learner's dictionaries | Find definitions, translations, and grammar explanations at Oxford learner's dictionaries. Retrieved from <https://www.oxfordlearnersdictionaries.com>
- Regoniel, P. (2015, January 5). Conceptual framework: A step-by-step guide on how to make one. Retrieved from <https://simplyeducate.me/2015/01/05/conceptual-framework-guide/>
- UN/Edifact glossary. (n.d.). Retrieved from https://www.unece.org/trade/untdid/texts/d300_s.htm

Sources for standard of requirements for definitions and concept systems.

- ISO 704 (2009). Technical Committee ISO/TC 37; Terminology and other language and content resources. Subcommittee SC 1; Principles and Methods. (2009). *Terminology Work – Principles and Methods*. Vernier, Switzerland: ISO.

Appendix 2: Frequencies

This Appendix shows in table 3 the frequencies of occurrence of data concepts found in 138 definitions of dimensions of data quality.

Table 3: Frequencies of data concepts in definitions of dimensions of data quality

Data concept	Frequency	%
Data	57	41,3%
Data values	8	5,8%
Attributes	5	3,6%
Format	4	2,9%
Data value	3	2,2%
Entity	3	2,2%
Values	3	2,2%
View	3	2,2%
Data pattern	2	1,4%
Dataset	2	1,4%
Entities	2	1,4%
Formats	2	1,4%
Information architecture or database	2	1,4%
Statistics	2	1,4%
Sub total	98	71,0%
Classification detail	1	
Components of the view	1	
Context	1	
Data accuracy	1	
Data availability	1	
Data content	1	
Data fields	1	
Data patterns in a time series	1	
Data presentation	1	
Data processing process	1	
Data, presentation, media, documentation	1	
Data/Redundant or distributed databases	1	
Database	1	
Date	1	
Datum	1	
Datum (e, a, v)	1	
Definitions	1	
Documentation	1	
Domain	1	
Entity types, attributes, domains	1	
Estimate	1	

Data concept	Frequency	%
Information presentation	1	
Initial estimated value	1	
Occurrences of data	1	
Physical instances of data	1	
Population	1	
Release time of data	1	
Representation of things	1	
Representations of a thing	1	
Statistic	1	
Statistical information	1	
Statistical output	1	
Subject data	1	
System or data	1	
Things	1	
Two datasets	1	
Units	1	
Usage of Recording Media	1	
Amount of data	1	
Capacity to change	1	
Total number of definitions	138	
Total number of data concepts	57	

15 data concepts occur twice or more in 138 definitions. 40 data concepts occur only once in a definition.

Appendix 3: Criteria for data concepts and data concept definitions

5.3 Criteria for data concepts

Criteria for data concepts are drawn up in table 4.

Table 4: Criteria for data concepts

Nr	Criterion
1	Relevant
2	Common
3	Unambiguous
4	Generic

5.4 Criteria for data concept definitions

Criteria for data concept definitions are drawn up in table 5.

Table 5: Criteria for definitions of data concepts

Code	Section in ISO 704	Criterion
A	6.2	Preferably formulated ' intensional ', not as an exhaustive list of possibilities.
B	6.2.1	Refers to concepts higher in the hierarchy .
C	6.3.1	The word ' is ' can be put in front of it.
D	6.3.2	Describes the associations with other concepts within the same concept system if possible
E	6.3.3	Is as short as possible and as complex as necessary.
F	6.3.4	Can replace the term in a sentence (substitution).
G	6.4.1	Is not circular .
H	6.4.2	Is not too wide and not too narrow .
I	6.4.3	Is formulated positive (not negative)
J	6.5	There are no additional comments in the definition. Comments occur in a Note.
<i>K-M</i>		<i>Not used here</i>
N		Is readable, unambiguous and written in proper language .
O		Corresponds with a generally accepted understanding of the meaning of the concept.

Appendix 4: Concepts and their definitions

This appendix is a compilation of definitions of **(data) concepts**. A selection is made of (data) concepts and their definitions that meet the criteria.

Caption

The column 'Source' contains the citation to a reference. If the column has no value, the authors of the report have formulated the definition whether or not derived from definitions from other sources.

The column 'Criteria not met' indicates by a letter which criteria for data concepts or data concept definitions are not met.

The column 'Frequency' means the number of times a term is used in a set of 138 definitions of data quality dimensions.

Preferred data concepts and definitions are marked blue

Each word that appears in **bold** in the definition of a term is a term defined elsewhere in this appendix.

1. Attribute

1.1 Definitions

Nr	Concept	Definition	Source	Criteria not met
1	Data attribute (1)	An inherent fact, property, or characteristic describing an entity or object; the logical representation of a physical field or relational table column. A given attribute has the same format, interpretations, and domain for all occurrences of an entity. Attributes may contain adjective valued (red, round, active, etc.)	Early 2011	J
2	Data attribute (2)	A unit of data for which the definitions, identifications, representation, and permissible values are specified by means of a set of characteristics.	Early 2011	E
3	Data attribute (3)	A representation of a data characteristic variation in the logical and physical data model. A data model may or may not be atomic.	Early 2011	J
4	Attribute	A characteristic of an object or entity.	ISO 11197	None
5	Attribute	An inherent characteristic, an accidental quality, an object closely associated with or belonging to a specific person, place, or office; a word ascribing a quality. (Bracket 2011)	In: Early 2011	B, E, J, O
6	Attribute	A piece of information that describes a Data Entity or Dictionary Entity.	ISO 21961	None
7	Attribute	A characteristic of an entity type about which the organisation wishes to hold information.	-	None

1.2 Analogies and frequencies

Analogy	Criteria not met	Frequency
Column	4	0
Data attribute	None	0
Data element	None	0
Data field	4	1
Field	4	0
Variable	2, 3	0
Attribute	None	5

1.3 Relationships

- Distinguishes **entity type**.
- Is specified by its name, definition, classification, and **format**.

1.4 Notes

- Early (2011) describes three definitions of data attributes without citing the source.
- Entity type is a population of entity instances which conform to the same data definition or schema, synonymous with object type or class. An entity type represents a class of objects in the universe of discourse, their world represented in a data model. They may be persons, places, things, abstract concepts, events, etc. of interest to the enterprise. (Everest 2010, cited in Early 2011)
- Example: Name is a data attribute of entity type Person.
- Five categories (subtypes) of attributes can be distinguished: 1) descriptive, 2) classifying, 3) quantitative, 4) time related, and 5) location related.

2. Characteristic

2.1 Definitions

Nr	Concept	Definition	Source	Criteria not met
1	Characteristic	Abstraction of a property	ISO 1087	None
2	Characteristic	Distinguishing feature	ISO 9000	None

2.2 Analogies

Analogy	Frequency	Criteria not met
Characteristic	0	None

2.3 Relationships

-

2.4 Notes

- Example: Having a cable for connecting with a computer' as a characteristic of the concept 'cord mouse'. (ISO 1087)

3. Composition of a dataset

3.1 Definitions

Nr	Concept	Definition	Source	Criteria not met
1	Composition	The nature of something's ingredients or constituents.	Lexico 2020	None
2	Composition	The way in which a whole or mixture is made up.	Lexico 2020	None
3	Composition of a dataset	The way in which a dataset is made up.	-	None

3.2 Analogies and frequencies

Analogy	Criteria not met	Frequency
Composition of a dataset	None	0

3.3 Relationships

-

3.4 Notes

The composition of a dataset can be unambiguously described by their:

1. Entity types and their definitions, e.g., person.
2. Data attributes and their definitions, e.g., name, city.
3. Classification schemes used in the dataset, e.g., a scheme for gender.
4. Population of entities, e.g., all persons living in London.
5. Period or point in time, e.g., all persons living in London in 2019 or on 31 December 2019.

4. Concept

4.1 Definitions

Nr	Concept	Definition	Source	Criteria not met
1	Concept	A unit of thought constituted through abstraction on the basis of characteristics common to a set of objects. [ISO 1087]	In: ISO 111079	None
2	Concept	Unit of knowledge created by a unique combination of characteristics	ISO 1087	None

4.2 Analogies

Analogy	Criteria not met	Frequency
Concept	None	Not applicable

4.3 Relationships

-

4.4 Notes

- Concepts are organized into concept systems. (ISO 704)
- Concepts are represented or expressed in language by designations or by definitions. (ISO 704)

5. Concept system

5.1 Definitions

Nr	Concept	Definition	Source	Criteria not met
1	Concept system	A set of concepts structured according to the relations among them.	ISO 704	None

5.2 Analogies and frequencies

Analogy	Criteria not met	Frequency
Concept system	None	Not applicable

5.3 Relationships

-

5.4 Notes

- Concept systems are a central element in terminology work and terminological research since concepts are not independent, discrete elements but belong to networks of other concepts, which are related to them in different ways.
- In order to understand even a single concept, one has to acquire knowledge about various other concepts in the same concept system before being able to define the concept reliably.
- Terminology work normally restricts itself to concept systems that structure the knowledge that specialists of a field share. Finding out and structuring these concept systems are seen as a prerequisite for achieving clarity in concepts and terms of special fields.

(Humbley et al. 2018)

6. Data

6.1 Definitions

Nr	Concept	Definition	Source	Criteria not met
1	Data	Facts represented as text, numbers, images, sound or video. Data is the raw material used to represent information, of from which information can be derived. (Everest 2010)	In: Early 2011	J
2	Data	The individual facts that are out of context and have no meaning by themselves. They are often referred to as raw data, such as 123.45. Data have historically been defined as plural; datum is the singular form. (Bracket 2011).	In: Early 2011	J
3	Data	A representation of facts, concepts, or instructions in a formalized manner, suitable for communication, interpretation, or processing by humans or by automatic means. (ISO 2382-4).	In: ISO 11179	None

6.2 Analogies

Analogy	Criteria not met	Frequency
Facts	2	0
Data	None	57

6.3 Relationships

-

6.4 Notes

-

7. Data category

7.1 Definitions

Nr	Concept	Definition	Source	Criteria not met
1	Data categories	Groupings of data with common characteristics or features.	McGilvray	E
2	Data category	A classification of data according to the purpose for which it is used.	-	-

7.2 Analogies and frequencies

Analogy	Criteria not met	Frequency
Data category	None	0

7.3 Notes

- Major data categories are master data, transactional data, reference data and metadata. Additional categories are historical data and temporary data.

8. Data concept

8.1 Definitions

Nr	Concept	Definition	Source	Criteria not met
2	Data concept	A concept related to data.	-	None

8.2 Analogies

Analogy	Criteria not met	Frequency
Data concept	None	Not applicable

8.3 Relationships

- Has associated dimensions.

8.4 Notes

- The term data concept is introduced to avoid the use of the more general term concept.

9. Data element

9.1 Definitions

Nr	Concept	Definition	Source	Criteria not met
1	Data element	See data attribute	Early 2011	-
2	Data element	A logical, identifiable unit of data that forms the basic organisational component in a database.	Jonker 2020	None
3	Data element	A unit of data which, in a certain context, is considered indivisible. [ISO 2382/4]	In: UN/Edifact	None
4	Data element	A unit of data for which the identification, description and value representation have been specified. [ISO 9735]	In: UN/Edifact	None
5	Data element	See data attribute	Early 2011	Not applicable
6	Data element	A unit of data for which the definition, identification, representation, and permissible values are specified by means of a set of attributes .	ISO 11179	None

9.2 Analogies

Analogy	Criteria not met	Frequency
Attribute		

9.3 Relationships

-

9.4 Note

- A data element is a special member of the broad notion called data. In general, data is a representation of facts, ideas, or instructions. Data is collected, organized, recorded, processed, and stored in a retrievable form. Data must also be suitable for communication, interpretation, or processing by human or automated means.
- There are many constructs used to organize data. There are data composites, entities, files, object classes, objects, records, relations, relationships, rows, segments, subject areas, tables, and tuples. None of these are analogous to data elements but may include or be supported by some database implementation or logical modelling equivalent of data elements.
- Bytes and bits are also components of data. Although they may be used to record data elements in an electronic medium, they do not correspond to data elements. In a database, a data element may be implemented as a field or column. In Chen's ER data model, it is an **attribute**. A data element then is a single unit of data that in a certain context is considered indivisible. It is a unit of data representing a single fact about a type of object (object class) in the natural world. (For example, a one-character code with allowed values of "M" or "S" representing the marital status attribute of an "employee" object class.) It cannot be decomposed into more fundamental segments of information that have useful meanings within the scope of its application. Data elements are thus defined as relevant to the user within the user's universe of discourse. Data elements are electronic or written representations of the properties of natural-world object classes.

(ISO 11179)

- Will not be used in our data concept system.

10. Data element specification

10.1 Definitions

Nr	Concept	Definition	Source	Criteria not met
1	Data element specification	The record of the set of attributes forming the definition, identification, representation and permissible values of a data element .	-	-

10.2 Analogies and frequencies

Analogy	Criteria not met	Frequency
Data element specification	None	0

10.3 Relationships with other data concepts

- Is a specification of an **data element** or **attribute**
- Specifies **data item**
- Includes specification of **value domain**

10.4 Notes

- The representation (of a data element) describes how the data are represented, i.e. the combination of a value domain, datatype, and, if necessary, a unit of measure or a character set. (ISO 11179 section 6)
- Will not be used in our data concept system.

11. Data item

11.1 Definitions

Nr	Concept	Definition	Source	Criteria not met
1	Data item	An individual field in a data record, referred to as a column in a relational database. A data item represents a data attribute, subject to adjustments made during formal data denormalization. Bracket (2011)	In: Early 2011	J
2	Data item	One occurrence of a data element .	ISO 11179	None
3	Data item	One occurrence of an attribute.	-	None

11.2 Analogies and frequencies

Analogy	Criteria not met	Frequency
Cell	4	0
Data item	None	0

11.3 Relationships with other data concepts

- Contains **data value**

11.4 Notes

-

12. Data pattern

12.1 Definitions

Nr	Concept	Definition	Source	Criteria not met
1	Data pattern	A series of data that repeats in a recognizable way.	Investopedia	None

12.2 Analogies and frequencies

Analogy	Criteria not met	Frequency
Data pattern	None	2

12.3 Notes

-

13. Data lineage

13.1 Definitions

Nr	Concept	Definition	Source	Criteria not met
1	Data lineage	A description of the pathway from the data source to their current location and the alterations made to the data along that pathway. (Brackett 2011)	In: Early 2011	
2	Data lineage	Metadata that identifies the sources of data and the transformation through which it has passed up to the point of consumption.	-	None

13.2 Analogies and frequencies

Analogy	Frequency	Criteria not met
Data lineage	0	None

13.3 Relationships with other data concepts

-

13.4 Notes

-

14. Data file

14.1 Definitions

Nr	Concept	Definition	Source	Criteria not met
1	File	A collection of information either on paper, or electronically in the form of data fields (or more complex structures) which describe a set of entities possessing some common characteristics or attributes; a collection of zero or more records which may have an arbitrarily complex structure (flat, hierarchical, etc.).	Early 2011	A, E, J
2	File	A collection of data, programs, etc. stored in a computer's memory or on a storage device under a single identifying name.	Oxford 2020	E
3	File	Information stored on a computer as one unit with one name.	Cambridge 2020	
4	Data file	A physical file of data that exists in a Database Management System , such as a computer file, or outside a database management system, such as a manual file. It is referred to as a table in a relational database. A data file generally represents a data entity, subject to adjustments made during formal data denormalization. (Brackett 2011)	In: Early 2011	E, J
5	Data file	Data stored on a computer as one unit with one name.	Cambridge 2020	

14.2 Analogies and frequencies

Analogy	Criteria not met	Frequency
Data file	None	2

14.3 Relationships with other data concepts

- A data file is a subtype of a file.
- A data file is part of a dataset.

14.4 Notes

-

15. Data value

15.1 Definitions

Nr	Concept	Definition	Source	Criteria not met
1	Data value	The specific representation of a value for an attribute as of a point in time.	Early 2011	E
2	Data value	An element of a value domain.	ISO 11179	None
3	Attribute value	A value associated with an attribute instance.	ISO 21961	None
4	Data value	The value of a data item .	-	None

15.2 Analogies and frequencies

Analogy	Criteria not met	Frequency
Attribute value	None	0
Value	None	3
Data value	None	3

15.3 Relationships with other data concepts

- Is contained in **data item**
- Forms part of **record**
- Is within **value domain**
- Represents a **property** of an **object**

15.4 Notes

-

16. Data value, initial

16.1 Definitions

Nr	Concept	Definition	Source	Criteria not met
5	Initial data value	A provisional data value that will be updated by a more accurate data value .	-	None

16.2 Analogies and frequencies

Analogy	Criteria not met	Frequency
Initial data value	None	0

16.3 Relationships with other data concepts

- Specification of a data value.

16.4 Notes

- The counterpart of an initial data value is a final data value.

17. Database

17.1 Definitions

Nr	Concept	Definition	Source	Criteria not met
1	Dataset	An organized collection of data stored in a structured way to enable rapid search and retrieval by a computer.	Early 2011	B
2	Database	a collection of related data typically organized in a computerized record keeping system that is part of a system whose purpose is to maintain the data and information derived from it so it can be made available for use.	ISO 22745-11 (ASTM)	B, E
3	Database	A collection of data organized according to a conceptual structure describing the characteristics of these data and the relationships among their corresponding entities, supporting one or more application areas.	ISO 22745-11 (ANSI)	B, E
4	Dataset	A dataset organized according to a structure describing the entity types , the attributes and the relationships between the entity types .	-	None

17.2 Analogies

Analogy	Criteria not met	Frequency
Database	None	0

17.3 Relationships with other data concepts

Subtype of a **dataset**.

17.4 Note

-

18. Dataset

18.1 Definitions

Nr	Concept	Definition	Source	Criteria not met
1	Dataset	Any organized collection of data .	Early 2011	None

18.2 Analogies

Analogy	Criteria not met	Frequency
File	4	5
Table	4	0
Dataset	None	1

18.3 Relationships with other data concepts

Is composed of **records**.

18.4 Note

-

19. Dataset availability

19.1 Definitions

Nr	Concept	Definition	Source	Criteria not met
1	Dataset availability	The degree to which a dataset can be consulted or retrieved by data consumers or processes.	-	None

19.2 Analogies

Analogy	Criteria not met	Frequency
Dataset availability	None	0

19.3 Relationships with other data concepts

Dataset availability is a characteristic of a dataset.

19.4 Note

Dataset availability can be substituted by dataset release or dataset delivery. These terms are equivalent.

20. Dataset composition

20.1 Definitions

Nr	Concept	Definition	Source	Criteria not met
1	Dataset composition	The way in which a dataset is made up.	-	None

20.2 Analogies

Analogy	Criteria not met	Frequency
Dataset composition	None	0

20.3 Relationships with other data concepts

-

20.4 Note

- The term dataset can be substituted by the term data file.

21. Definition

21.1 Definitions

Nr	Concept	Definition	Source	Criteria not met
1	Definition	A word or phrase expressing the essential nature of a person or thing or class of persons or things	ISO 11179	None
2	Definition	A statement of the meaning of a word or word group or a sign or symbol	Merriam-Webster 2020	None
3	Definition	An explanation of the meaning of a word or phrase, especially in a dictionary	Oxford 2020	None
4	Definition	Representation of a concept by an expression that describes it and differentiates it from related concepts	ISO 1087	None

21.2 Analogies and frequencies

Analogy	Frequency	Criteria not met
Definition	Not applicable	None

21.3 Relationships with other data concepts

-

21.4 Notes

(ISO 11179)

- A word or phrase expressing the essential nature of a person or thing or class of persons or things: an answer to the question “what is x?” or “what is an x?”
- Statement that expresses the essential nature of a data element and permits its differentiation from all other data elements.

22. Dimension

22.1 Definitions

Nr	Concept	Definition	Source	Criteria not met
1	Dimension	Measurable characteristic.	DAMA 2017	None

22.2 Analogies and frequencies

Analogy	Frequency	Criteria not met
Characteristic	0	None
Feature	0	2
Dimension	0	None

22.3 Relationships with other data concepts

Is associated with a **data concept.**

22.4 Notes

-

23. Entity type

23.1 Definitions

Nr	Concept	Definition	Source	Criteria not met
1	Entity type	A population of entity instances which conform to the same data definition or schema, often synonymous with object type or class. An entity type represents a class of object in the users' universe of discourse, their world represented in a data model. They may be persons, places, things, abstract concepts, events, etc. of interest to the enterprise. (Everest 2010)	In: Early 2011	B, E, J
2	Entity type	None	ISO 11179	-
3	Entity type	A classification of the types of real-world objects (such as person, place, thing, concept, or events of interest to the enterprise) that have common characteristics. Sometimes the term <i>entity</i> is used as a short name.	English 1999	J
4	Data entity	A concept that can, or does, take on one or more values.	ISO 21961	H, N
5	Entity type	A thing of significance about which the organisation wishes to hold information	Hay 2013	

23.2 Analogies and frequencies

Analogy	Frequency	Criteria not met
Class	0	4
Data entity	0	3
Object class	0	4
Object type	0	4
Entity type	0	None

23.3 Relationships with other data concepts

- Is distinguished by **attributes**
- Describes **objects**

23.4 Notes

-

24. Format

24.1 Definitions

Nr	Concept	Definition	Source	Criteria not met
1	Format	The specifications for layout or display of information, such as in a document or on a disk.	Early 2011	J
2	Representation of a data element	Describes how the data are represented, i.e. the combination of a value domain, datatype, and, if necessary, a unit of measure or a character set.	ISO 11179	C, J
3	Format	A combination of value domain , datatype, unit of measure and character set.	-	A, B

24.2 Analogies and frequencies

Analogy	Criteria not met	Frequency
Representation	3	2
Format	None	2

24.3 Relationships with other data concepts

- Is part of a specification of a data element or an **attribute**.

24.4 Notes

-

25. Master data

25.1 Definitions

Nr	Concept	Definition	Source	Criteria not met
1	Master Data	Data that provides the context for business activity data in the form of common and abstract concepts that relate to the activity.	Early 2011	B
2	Master Data	Data held by an organization which describe the entities that are both independent and fundamental for that organization, and <u>that</u> it needs to reference in order to perform its transactions.	ISO 22745-2:2010	B, E
3	Master Data	Data held by an organization which describe the entities that are both independent and fundamental for that organization, and <u>which</u> it needs to reference in order to perform its transactions.	ISO 8000-2:2017	B, E
4	Master data	Describe the people, places, and things that are involved in an organization's business.	McGilvray 2008	C
5	Master Data	Data held by an organization which describe object types that it needs to reference in order to perform its transactions.	-	None

25.2 Analogies

Analogy	Criteria not met	Frequency
Master Data	None	0

25.3 Relationships with other concepts

- Is an instance of **data category**.

25.4 Notes

- Examples include people (e.g., customers, employees, vendors, suppliers), places (e.g., locations, sales territories, offices), and things (e.g., accounts, products, assets, document sets).
- Because these data tend to be used by multiple business processes and IT systems, standardizing master data formats and synchronizing values are critical for successful system integration.
- Master data tend to be grouped into master records, which may include associated reference data. An example of associated reference data is a state field within an address in a customer master record.

(McGilvray 2008)

26. Metadata

26.1 Definitions

Nr	Concept	Definition	Source	Criteria not met
1	Meta-data	Literally, “data about data”; data that defines and describes the characteristics of other data, used to improve both business and technical understanding of data and data-related processes. Because the term ‘metadata’ is a trademark of The Metadata Company, DAMA specifically uses the term meta-data.	Early 2011	J
2	Metadata	Label, describe or characterize other data and make it easier to retrieve, interpret, or use information.	McGilvray	B, C, E, F
3	Metadata	Data that defines and describes other data .	ISO 11179	None

26.2 Analogies and frequencies

Analogy	Frequency	Criteria not met
Metadata	0	None

26.3 Relationships with other data concepts

-

26.4 Notes

-

27. Object

27.1 Definitions

Nr	Concept	Definition	Source	Criteria not met
1	Object	In the real world, a person, place, thing or concept.	Early 2011	None
2	Object	Any part of the conceivable or perceivable world.	ISO 1087	None
3	Object	Any part of the conceivable or perceivable world. (ISO 1087)	In: ISO 11179	None
4	Entity	Any concrete or abstract thing, that exists, did exist, or might exist, including associations among these things, e.g., a person, object, event, idea, process, etc.	Early 2011	J
5	Entity	Any concrete or abstract thing of interest, including associations among things. (ISO/IEC 2382)	In: ISO 11179	E
6	Object	Anything perceivable or conceivable.	ISO 9000	None

27.2 Analogies and frequencies

Analogy	Frequency	Criteria not met
Entity	2	None
Object	0	None

27.3 Relationships with other data concepts

- Is described by **entity type**
- Is characterised by **properties**
- Is represented by **records**

27.4 Notes

- Objects can be material (e.g. engine, sheet of paper, diamond), immaterial (e.g. conversion ratio, project plan) or imagined (e.g. unicorn, scientific hypothesis). (ISO1087).

28. Object type

28.1 Definitions

Nr	Concept	Definition	Source	Criteria not met
1	Object class	A set of objects. A set of ideas, abstractions, or things in the real world that can be identified with explicit boundaries and meaning and whose properties and behaviour follow the same rules.	ISO 11179	E
2	Object type	A set of objects with explicit boundaries and meaning and whose properties and behaviour follow the same rule.	-	None

28.2 Analogies and frequencies

Analogy	Frequency	Criteria not met
Object class	0	4
Object type	0	None

28.3 Relationships with other data concepts

-

28.4 Notes

- Will not be used in our data concept system.

29. Property

29.1 Definitions

Nr	Concept	Definition	Source	Criteria not met
1	Property	A peculiarity common to all members of an object class.	ISO 11179	None
2	Property	A feature of an object .	ISO 1087	None

29.2 Analogies and frequencies

Analogy	Frequency	Criteria not met
Property	0	None

29.3 Relationships with other data concepts

- Characterises an **object**
- Is recorded by a **data value**
- Actually, has a **real value**

29.4 Notes

- Example: 'Being made of wood' as a property of a given 'table'. (ISO 1087).
- Refers to an (individual) object and not to an object type.

30. Transactional data

30.1 Definitions

Nr	Concept	Definition	Source	Criteria not met
1	Transactional data	Describe an internal or external event or transaction that takes place as an organization conducts its business.	McGilvrey 2008	C, E, G
2	Transactional data	Data that describes an event that takes place as an organization conducts its business.	-	-

30.2 Analogies and frequencies

Analogy	Frequency	Criteria not met
Transactional data	0	None

30.3 Relationships with other data concepts

- Is an instance of **data category**.

30.4 Notes

- Examples include sales orders, invoices, purchase orders, shipping documents, passport applications, credit card payments, and insurance claims.
- These data are typically grouped into transactional records, which include associated master and reference data.

(McGilvrey 2008)

31. Real value

31.1 Definitions

Nr	Concept	Definition	Source	Criteria not met
1	Real value	The real-life value of a property of an object	-	None

31.2 Analogies and frequencies

Analogy	Frequency	Criteria not met
True value	0	None
Actual value	0	2
Real value	0	None

31.3 Relationships with other data concepts

Expresses an instance of a **property**.

31.4 Notes

-

32. Record

32.1 Definitions

Nr	Concept	Definition	Source	Criteria not met
1	Data record	A physical grouping of data items that are stored in or retrieved from a data file. It is referred to as a row or tuple in a relational database. A record represents a data instance. (Bracket 2011)	In: Early 2011	E, J
2	Record	A logically related set of data values that represent a (real-world) object .	-	None

32.2 Analogies and frequencies

Analogy	Criteria not met	Frequency
Data record	None	0
Entity	3	0
Member	2	0
Object	3	0
Row	4	0
Unit	2	0
Record	None	9

32.3 Relationships with other data concepts

- Forms part of **data file**
- Is composed of **data values**

32.4 Notes

-

33. Reference data

33.1 Definitions

Nr	Concept	Definition	Source	Criteria not met
1	Reference data	Sets of values or classification schemas that are referred to by systems, applications, data stores, processes, and reports, as well as by transactional and master records	McGivrey 2008	B
2	Reference data	Generally, any data used to organize or categorize other data, or for relating data to information both within and beyond the boundaries of the enterprise. Usually consists of codes and descriptions or definitions.	Early 2011	E, J
2	Reference data	Data used to categorize other data.	-	-

33.2 Analogies and frequencies

Analogy	Criteria not met	Frequency
Reference data	None	9

33.3 Relationships with other data concepts

- Is an instance of **data category**.

33.4 Notes

- Examples include lists of valid values, code lists, status codes, state abbreviations, demographic fields, flags, product types, gender, chart of accounts, and product hierarchy.
- Standardized reference data are key to data integration and interoperability and facilitate the sharing and reporting of information.
- Reference data may be used to differentiate one type of record from another for categorization and analysis, or they may be a significant fact such as country, which appears within a larger information set such as address.
- Organizations often create internal reference data to characterize or standardize their own information.
- Reference data sets are also defined by external groups, such as government or regulatory bodies, to be used by multiple organizations. For example, currency codes are defined and maintained by ISO.

(McGivrey 2008)

34. Statistical output

34.1 Definitions

Nr	Concept	Definition	Source	Criteria not met
1	Statistical output	Output from a statistical process.	-	None

34.2 Analogies and frequencies

Analogy	Frequency	Criteria not met
Statistical output	0	None

34.3 Relationships with other data concepts

- Is an instance of **data category**.

34.4 Notes

-

35. Value domain

35.1 Definitions

Nr	Concept	Definition	Source	Criteria not met
1	Value domain	A set of permissible values.	ISO 11197	
2	Value domain	A set of permissible values of an attribute .	-	None

35.2 Analogies and frequencies

Analogy	Criteria not met	Frequency
Value domain	None	0

35.3 Relationships with other data concepts

- Includes **data value**

35.4 Notes

-

Version history

Version	Date	Description of the modification	Author
1.0.p1	28 April 2020	First draft	Peter
1.0.p2	4 May 2020	Comment Andrew processed (first round).	Peter
1.0.p3	9 May 2020	Comment Andrew processed (second round)	Peter
1.0.p5	19 May 2020	Edits applied to Chapter 1. Comment Andrew processed (third round)	Andrew and Peter
1.0.p7	1 June 2020	Edits applied. Changes processed. Relationships added for each data concept.	Andrew, Peter
1.0.p8	10 June 2020	Comment reviewers processed. Figures adapted. Object type, data element and data element specification are removed from the data concept system. Ready for review by Andre.	Peter, Andrew
1.0.p9	14 June 2020	Edits applied	Andrew
1.0.p10	4 August 2020	Data Lineage added.	Peter
1.0	5 August 2020	Minor edits. Final version	Andrew
1.1	20 August 2020	Three data concepts added.	Peter
1.2	3 Sept 202	Small edits	Peter

Active distribution per version	
Version	Distribution
1.0.p1-p6	Dropbox
1.0.p7	Reviewers
1.0.p8	Dropbox
1.0.p9	Dropbox, Reviewers.
1.0.p10	Dropbox, Wim Stolk
1.0	Dropbox, DAMA-NL
1.1	Dropbox, DAMA-NL
1.2	Dropbox, Website DAMA-NL