

# How to Select the Right Dimensions of Data Quality

Includes 60 dimensions of data quality  
and their standardized definitions



## Colophon

### Authors:

Andrew Black (Van Nederpelt & Black)

Peter van Nederpelt (Van Nederpelt & Black)

Version: See Version History

Printed of 5-9-2020 11:23

© DAMA NL Foundation. All rights reserved.

<b>Table of contents</b>	
<b>Tables</b>	<b>5</b>
<b>Figures</b>	<b>5</b>
<b>1. Introduction</b>	<b>6</b>
1.1 The importance of data quality	6
1.2 Purpose of this document	6
1.3 Audience of this document	6
1.4 How did the document come about?	6
1.5 Release policy	6
1.6 Reading guideline	6
<b>2. Definitions and examples</b>	<b>7</b>
2.1 What is data quality?	7
2.2 What is a dimension?	7
2.3 What is a data concept?	7
2.4 What is a dimension of a data concept?	8
2.5 What is a requirement?	8
<b>3. How to select the right dimensions of data quality?</b>	<b>10</b>
3.1 Procedure to improve control or improve data quality	10
3.2 Step 1: Determine which dimensions of data quality are important for the present data category	10
3.3 Step 2: Determine whether a dimension contributes sufficiently to a higher objective	10
3.4 Step 3: Check whether all relevant dimensions have been identified	11
3.5 Step 4: Establish indicators and associated measurements methods for the selected dimensions	11
<b>Appendix 1: Dimensions of data quality</b>	<b>12</b>
<b>Appendix 2: Combinations of dimensions and data categories</b>	<b>17</b>
<b>Appendix 3: Elaborated dimensions of data quality</b>	<b>20</b>
Accuracy	21
Availability	21
Clarity	22
Completeness (1)	22
Completeness (2)	22
Consistency	23
Currency	23
Punctuality	24
Timeliness	24

Traceability	25
Uniqueness	25
Validity	26
<b>Appendix 4: Definitions of concepts and data concepts</b>	<b>27</b>
<b>Appendix 5: Diagrams</b>	<b>30</b>
<b>Appendix 6: Sources</b>	<b>32</b>
Sources of definitions of quality dimension	32
Other sources	32
<b>Version history</b>	<b>34</b>

## **Tables**

**Table 1:** Examples of combinations of dimensions and data concepts

**Table 2:** Examples of requirements of data quality

**Table 3:** Actions to control improve data quality

**Table 4:** Definitions of dimensions of data quality

**Table 5:** Combinations of dimensions and data categories

**Table 6:** Definitions of concepts

## **Figures**

**Figure 1:** Data concepts in a data model

**Figure 2:** Relationship between data concepts and dimensions

**Figure 3:** Artists impression of the real world and data world

## **1. Introduction**

### **1.1 The importance of data quality**

Data plays an increasingly important useful role in our society. Dependence on data for many activities and processes is increasing. Quality of data is therefore of growing importance and should be managed.

### **1.2 Purpose of this document**

The purpose of this document is to present an approach to selecting the dimensions of data quality that best apply to a specific situation. This is the first step to control or improve data quality.

### **1.3 Audience of this document**

The report is meant for everyone who is involved in management of data quality, particularly those preparing to apply data quality dimensions in practice.

### **1.4 How did the document come about?**

This document is an initiative of the Data Quality working group of DAMA-NL. This working group drew up a research paper about dimensions of data quality (Black, Van Nederpelt, 2020). Subsequently, the present report has been derived from this paper. Finally, it was submitted to the DAMA community for comment and published.

### **1.5 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](mailto:info@dama-nl.org) or the authors [info@vannederpeltblack.nl](mailto:info@vannederpeltblack.nl).

### **1.6 Reading guideline**

Chapter 1 describes the purpose and use of this document.

Chapter 2 explains some key concepts.

Chapter 3 presents the steps to select the right dimensions of data quality as part of a procedure to control or improve data quality.

Appendix 1 shows all 60 dimensions of data quality and their definitions.

Appendix 2 presents logical combinations of dimensions of data quality and data categories.

Appendix 3 elaborates twelve common dimensions of data quality.

Appendix 4 defines concepts used in the report.

Appendix 5 shows diagrams of the data concept system.

Appendix 6 contains references.

## 2. Definitions and examples

This chapter presents some important concepts and their definitions.

### 2.1 What is data quality?

**Data quality** is the degree to which **dimensions of data concepts** meet **requirements**.

This definition is derived from the definition of quality of the well-known standard for quality management systems ISO 9001. This standard defines quality as the degree to which inherent characteristics of an object meet requirements (ISO 9000:2015).

To make the definition more specific for data, we replaced 'characteristic' by 'dimension' and 'object' by 'data concept'.

### 2.2 What is a dimension?

**Dimension** is a measurable feature of a data concept.

This definition is derived from the definition from ISO 9000. This standard defines a characteristic as a feature of an object.

The term dimension is used to make the connection to dimensions in the measurement of physical objects (e.g., length, width, height). Examples of dimensions are accuracy, completeness, and timeliness.

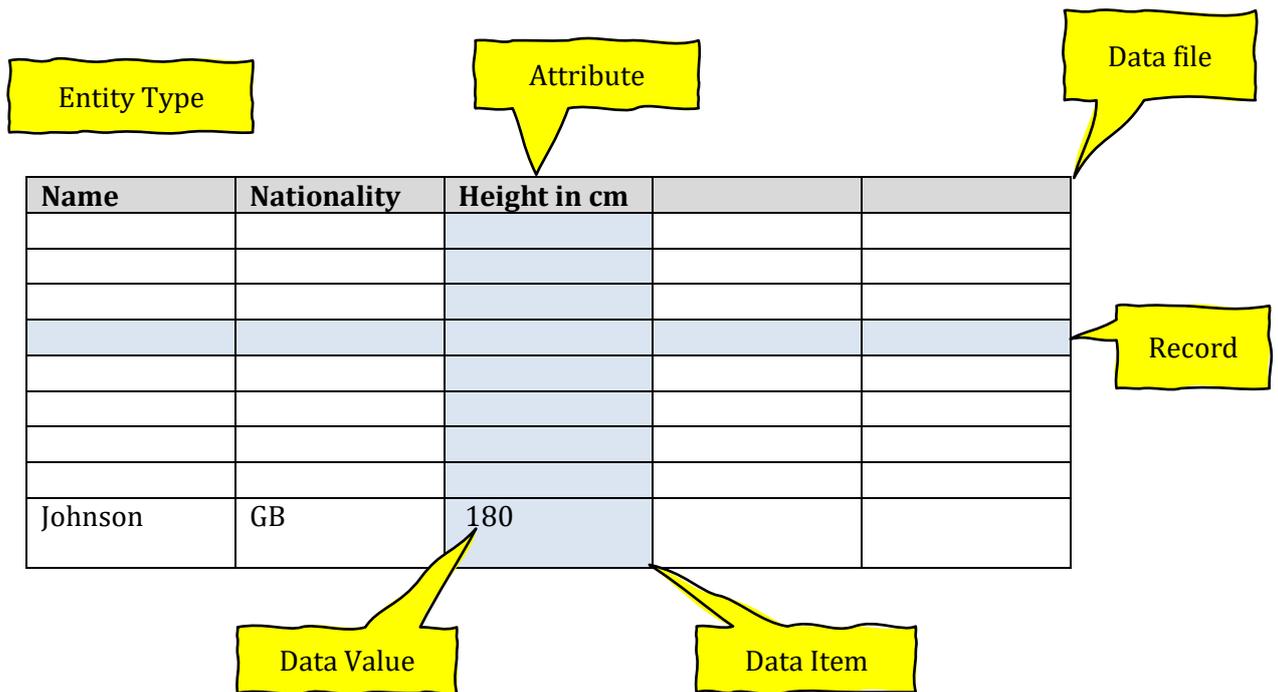
The term dimension in this context should not be confused with its use in the context of business intelligence where it refers to a category for summarizing or viewing data.

### 2.3 What is a data concept?

Definitions of the term data concept and the data concepts themselves can be found in Appendix 4.

Examples of **data concepts** are data files, records, attributes, and data values.

A subset of data concepts in the data world is shown in a data model in Figure 1. In Appendix 5, two other diagrams of data concepts are presented.



**Figure 1:** Data concepts in a data model

## 2.4 What is a dimension of a data concept?

A dimension of a data concept is a logical combination of a dimension and a data concept. Examples are combinations of dimensions and data concepts are presented in Table 1.

**Table 1:** Examples of combinations of dimensions and data concepts

Dimension	Data concept
Accuracy	Data values
Completeness	Records
Completeness	Data values
Referential integrity	Data files (tables)

See also Figure 2 in Appendix 5.

## 2.5 What is a requirement?

Requirement is defined in ISO 9000 as:

**Requirement** is a need or expectation that is stated, generally implied or obligatory.

Examples of requirements are shown in Table 2. In the context of data quality, requirements can be made specific by target values of indicators that are associated with dimensions of data quality.

**Table 2:** Examples of requirements of data quality

<b>Dimension</b>	<b>Data concept</b>	<b>Requirement</b>
Accuracy	Data values	The names in a customer file should be more than 96% correctly spelled.
Completeness	Records	The product file should contain 99,5% of the products that the company sells.
Referential integrity	Data files	All employees in the employee file should be linked to a department file.

It should be noted that requirements for dimensions of data quality are context dependant and should be established by the data producer or data consumer. You cannot state in general that the quality in all cases should be as high as possible because unnecessary costs may be incurred.

### 3. How to select the right dimensions of data quality?

This chapter describes how dimensions of data quality can be selected. These actions are part of a procedure to control or improve data quality. Therefore, we describe first this procedure briefly to put the selection actions into context.

#### 3.1 Procedure to improve control or improve data quality

Table 3 describes the actions to control or improve data quality.

**Table 3:** Actions to control or improve data quality

Step	Action
Dimensions of data quality	
1	Determine which dimensions of data quality are important for the present data category.
2	Determine whether a dimension contributes sufficiently to a higher objective.
3	Check whether all relevant dimensions have been identified.
Indicators	
4	Establish indicators and associated measurements methods for the selected dimensions.
5	Establish target values for these indicators.
6	Measure the value of these indicators and check whether they meet the target values.
Measures	
7	Take measures to improve the data quality.
Evaluation	
8	Evaluate the dimensions, the indicators, and the measures.

The actions about selecting dimensions of data quality (step 1 – 3) will be elaborated in the next paragraphs. Some attention is also paid to step 4 about indicators.

#### 3.2 Step 1: Determine which dimensions of data quality are important for the present data category

Determine the data category. Examples of data categories are master data, reference data, transactional data, registers, and statistical output.

Determine which dimensions are important for the data category. Table 5 in Appendix 2 indicates which dimensions are candidates for a specific data category.

#### 3.3 Step 2: Determine whether a dimension contributes sufficiently to a higher objective

Determine whether a dimension contributes sufficiently to a higher objective. The contribution must be large enough to select the dimension and measure the associated indicator.

The following are examples of possible objectives:

- A. Satisfaction of customers and other stakeholders
- B. Quality of the product or service delivered by the organisation

- C. Public confidence in the organisation
- D. Reputation of the organisation
- E. Interoperability between organisations
- F. The level of data quality management costs relative to the costs of emergency data quality repairs and the risks of fines due to non-compliance
- G. Efficiency of the processes of all partners in a data processing chain
- H. Compliance of the organisation with laws and regulations

**3.4 Step 3: Check whether all relevant dimensions have been identified**

Check whether no relevant dimensions have been forgotten. See Appendix 1 for 60 dimensions of data quality and their definitions.

**3.5 Step 4: Establish indicators and associated measurements methods for the selected dimensions**

Establish indicators for the selected dimensions. Appendix 3 shows possible indicators for some common indicators.

Establish a measurement method for each indicator.

## Appendix 1: Dimensions of data quality

This Appendix defines sixty dimensions of data quality. These dimensions can be found in various sources related to data management. Table 4 shows the dimensions and their definitions in alphabetical order.

Common dimensions are **marked blue** and elaborated in Appendix 3.

The following principles have been applied in compiling the definitions of the dimensions of data quality:

- The list has been made as complete as possible.
- Definitions that already exist have been used as much as possible.
- The definitions meet the requirements of ISO 704. This standard is about defining terms in general. For example, a definition should not be too long and should not contain examples.
- The definition always starts with 'the degree to which...'
- A dimension is always part of something. We call it a data concept (e.g. attribute, record, or data file).
- The data concepts together form a data concept system. These data concepts are also defined and visualised. See Appendix 4 and 5.
- Dimensions of data quality can be classified by data concept.

**Table 4:** Definitions of dimensions of data quality

Nr	Dimension	Data Concept	Definition
1.	Ability to represent null values	Format	The degree to which a format allows null values in an attribute.
2.	Access security	Datasets	The degree to which access to datasets is restricted.
3.	Accessibility	Data	The ease with which data can be consulted or retrieved,
4.	Accuracy	Data values	The degree of closeness of data values to real values.
5.	Appropriateness	Format	The degree to which the format is suitable for use.
6.	Availability	Data	The degree to which data can be consulted or retrieved by data consumers or a process.
7.	Clarity	Metadata	The ease with which data consumers can understand the metadata.
8.	Coherence	Composition of datasets	The degree to which datasets can be combined.
9.	Comparability of populations	Data values	The degree to which data values representing two populations have the same definition and are measured in the same way.
10.	Comparability over time	Data values	The degree to which data values over time have the same definition and are measured in the same way.

Nr	Dimension	Data Concept	Definition
11.	Completeness	Attributes	The degree to which all required attributes in the dataset are present.
12.		Records	The degree to which all required records in the dataset are present.
13.		Data files	The degree to which all required data files are present.
14.		Data values	The degree to which all required data values are present.
15.		Data values of an attribute	The degree to which all required data values of an attribute are present.
16.		Metadata	The degree to which the metadata are fully described.
17.	Compliance with laws, regulations, or standards	Data	The degree to which data is in accordance with laws, regulations, or standards.
18.		Composition of datasets	The degree to which the composition of datasets is in accordance with laws, regulations, or standards.
19.	Confidentiality	Data	The degree to which disclosure of data should be restricted to authorized data consumers.
20.	Consistency	Data values	The degree to which data values of two sets of attributes <ul style="list-style-type: none"> <li>▪ within a record,</li> <li>▪ within a data file,</li> <li>▪ between data files,</li> <li>▪ within a record at different points in time</li> </ul> comply with a rule.
21.		Data values of a set of attributes of a dataset at different points in time (temporal consistency)	The degree to which the data values of a set of attributes of a dataset at different points in time comply with a rule.
22.		Data values of two sets of attributes between datasets (across datasets)	The degree to which data values of two sets of attributes between datasets comply with a rule.
23.		Data values of two sets of attributes between records (cross record)	The degree to which data values of two sets of attributes between records comply with a rule.

<b>Nr</b>	<b>Dimension</b>	<b>Data Concept</b>	<b>Definition</b>
24.		Data values of two sets of attributes within a record (record level)	The degree to which data values of two sets of attributes within a record comply with a rule.
25.	Credibility	Data values	The degree to which data values are regarded as true and believable by data consumers.
26.	Currency	Data values	The degree to which data values are up to date.
27.	Equivalence	Attributes	The degree to which attributes stored in multiple datasets are conceptually equal.
28.	Granularity	Attributes	The degree to which a single characteristic is subdivided in attributes.
29.	Granularity	Records	The degree to which objects are aggregated to records.
30.	Integrity	Data values	The degree of absence of data value loss or corruption.
31.	Interpretability	Data	The degree to which data are in an appropriate language and units of measure.
32.	Latency	Data	The period of time between the point when the data is created and the point when it is available for use.
33.	Linkability	Data files	The degree to which records of one data file can be correctly coupled with records of another data file.
34.	Metadata compliance	Data values	The degree to which the data values are in accordance with their definition, format specification and value domain.
35.	Naturalness	Composition of datasets	The degree to which the composition of datasets is aligned with the real-world objects that they represent.
36.	Objectivity	Data values	The degree to which the data values are created in an unbiased manner.
37.	Obtainability	Data	The degree to which the data can be acquired.
38.	Plausibility	Data values	The degree to which data values match knowledge of the real world.

<b>Nr</b>	<b>Dimension</b>	<b>Data Concept</b>	<b>Definition</b>
39.	Portability	Data	The degree to which data can be installed, replaced, or moved from one system to another while preserving the existing quality.
40.	Portability	Format	The degree to which a format can be applied in a wide range of situations.
41.	Precision (1)	Data values	The degree of accuracy with which data value are recorded or classified.
42.	Precision (2)	Data values	The degree to which the error in data values spreads around zero (in statistics).
43.	Punctuality	Dataset availability	The degree to which the period between the actual and target point of time of availability of a dataset is appropriate.
44.	Reasonability	Data pattern	The degree to which a data pattern meets expectations.
45.	Recoverability	Datasets	The degree to which datasets are preserved in the event of incident.
46.	Redundancy	Data	The degree to which logically identical data are stored more than once.
47.	Referential integrity	Data files	The degree to which data values of the primary key of one data file and data values of the foreign key of another data file are equal.
48.	Relevance	Composition of datasets	The degree to which the composition of datasets meets the needs of the data consumer.
49.	Reliability	Initial data value	The closeness of the initial data value to the subsequent data value.
50.	Reproducibility	Dataset	The degree to which a dataset can be recreated with the same data values.
51.	Reputation	Data	The degree to which data are trusted or highly regarded in terms of their source or content.
52.	Retention period	Datasets	The period that datasets are available until they can or must be deleted.

<b>Nr</b>	<b>Dimension</b>	<b>Data Concept</b>	<b>Definition</b>
53.	Timeliness	Dataset availability	The degree to which the period between the time of creation of the real value and the time that the dataset is available is appropriate.
54.	Traceability	Data	The degree to which data lineage is available.
55.	Uniqueness	Objects	The degree to which objects (of the real world) occur only once as a record in a data file.
56.		Records	The degree to which records occur only once in a data file.
57.	Validity	Data values	The degree to which data values comply with rules.
58.	Value	Data	The degree to which data provide advantages from their use.
59.	Variety	Data	The degree to which data are available from different data sources.
60.	Volatility	Data values	The degree to which data values change over time.

Source: Black, A., Nederpelt, P. van. (2020). *Dimensions of Data Quality Dimensions*. Research paper. DAMA-NL.

## Appendix 2: Combinations of dimensions and data categories

Table 5 indicates which dimensions are candidates for selection in case of a specific data category.

The dimensions in the column statistical output are numbered because in the statistical domain these dimensions are usually presented in this sequence.

**Table 5:** Combinations of dimensions and data categories

Nr	Dimension	Data Concept	Data Category					Statistical Output
			Common	Master Data	Reference Data	Transactional Data	Registers	
1	Ability to represent null values	Format						
2	Access security	Datasets		X	X	X	X	
3	Accessibility	Data						9
4	Accuracy	Data values	X	X	X	X	X	2
5	Appropriateness	Format						
6	Availability	Data	X					
7	Clarity	Metadata	X	X	X	X	X	10
8	Coherence	Composition of datasets						6
9	Comparability of populations	Data values						8
10	Comparability over time	Data values						7
14	Completeness	Data values	X	X	X	X	X	
13	Completeness	Data files						
15	Completeness	Data values of an attribute						
12	Completeness	Records	X	X	X	X	X	
11	Completeness	Attributes						
16	Completeness	Metadata		X	X	X	X	
17	Compliance with laws, regulations, or standards	Data					X	
18	Compliance with laws, regulations, or standards	Composition of datasets					X	
19	Confidentiality	Data				X		
20	Consistency	Data values						

Nr	Dimension	Data Concept	Data Category					
			Common	Master Data	Reference Data	Transactional Data	Registers	Statistical Output
21	Consistency	Data values of a set of attributes of a dataset at different points in time (temporal consistency)						
22	Consistency	Data values of two sets of attributes between datasets (across datasets)						
23	Consistency	Data values of two sets of attributes between records (cross record)						
23	Consistency	Data values of two sets of attributes within a record (record level)						
25	Credibility	Data values						
26	Currency	Data values	X	X	X	X	X	
27	Equivalence	Attributes						
29	Granularity	Records						
28	Granularity	Attributes						
30	Integrity	Data values						
31	Interpretability	Data						
32	Latency	Data						
33	Linkability	Data files						
34	Metadata compliance	Data values						
35	Naturalness	Composition of datasets						
36	Objectivity	Data values						
37	Obtainability	Data						
38	Plausibility	Data values						
40	Portability	Format						
39	Portability	Data						
41	Precision (1)	Data values						
42	Precision (2)	Data values						
43	Punctuality	Dataset availability	X					5
44	Reasonability	Data pattern						
45	Recoverability	Datasets		X	X	X	X	
46	Redundancy	Data						
47	Referential integrity	Data files						
48	Relevance	Composition of datasets						1

Nr	Dimension	Data Concept	Data Category					Statistical Output
			Common	Master Data	Reference Data	Transactional Data	Registers	
49	Reliability	Initial data value						3
50	Reproducibility	Dataset						
51	Reputation	Data						
52	Retention period	Datasets				X		
53	Timeliness	Dataset availability	X					4
54	Traceability	Data	X			X		
56	Uniqueness	Records	X	X	X		X	
55	Uniqueness	Objects					X	
57	Validity	Data values	X			X	X	
58	Value	Data						
59	Variety	Data						
60	Volatility	Data values						

### **Appendix 3: Elaborated dimensions of data quality**

In this Appendix twelve common dimensions of data quality are elaborated.

1. Accuracy
2. Availability
3. Clarity
4. Completeness of records
5. Completeness of data values
6. Consistency
7. Currency
8. Punctuality
9. Timeliness
10. Traceability
11. Uniqueness
12. Validity

For each dimension, the following items are described:

- Title. Name of the dimension.
- Long title. Name of the dimension and the associated data concept.
- Synonyms
- Related. Dimensions that are dependent on or contributes to the dimension.
- Definition
- Indicators. Possible indicators.
- Examples. Descriptions of non-compliances with required data quality.
- Notes

## Accuracy

<b>Title</b>	Accuracy
<b>Long title</b>	Accuracy of data values
<b>Synonym</b>	Correctness of data values
<b>Related</b>	-
<b>Definition</b>	The degree of closeness of data values to real values.
<b>Indicators</b>	Percentage or number of inaccurate data values.
<b>Examples on non-compliances</b>	<ul style="list-style-type: none"> <li>▪ A house is located at number 120 but registered as number 12.</li> <li>▪ A person is called Janssen but registered as Jansen.</li> <li>▪ A farm has 7,321 chickens. It is registered as 7,321 while the unit of measurement is thousand. It should be registered as 7.</li> <li>▪ A product is located at A23 but according to the database its location is P76.</li> <li>▪ The number of unemployed people is estimated at 234.000. If the sample is not fully representative, there will be bias or systematic error. The size of the sample determines the variance or random error of the estimate.</li> </ul>
<b>Notes</b>	<ul style="list-style-type: none"> <li>▪ The data producer or consumer must define when he/she considers a data value as inaccurate and define criteria for inaccuracy.</li> <li>▪ The impact of an inaccuracy is different for each attribute.</li> <li>▪ Generally, accuracy will be measured for individual attributes, e.g., the accuracy of the product name.</li> </ul>

## Availability

<b>Title</b>	Availability
<b>Long title</b>	Availability of data
<b>Synonyms</b>	-
<b>Related</b>	Obtainability of data
<b>Definition</b>	The degree to which data can be consulted or retrieved by data consumers or a process.
<b>Indicators</b>	<ul style="list-style-type: none"> <li>▪ Yes or No</li> <li>▪ The effort it takes to make data available (hours)</li> </ul>
<b>Examples on non-compliances</b>	<ul style="list-style-type: none"> <li>▪ Data are not available because they are not processed yet such as the number of casualties of a recent incident.</li> <li>▪ Personal data are not available to the public.</li> <li>▪ Data are not available for reasons of competition.</li> <li>▪ Data are not available because they are confidential or secret.</li> <li>▪ Data are not available because they not archived in a professional manner.</li> </ul>
<b>Notes</b>	Data can be partly available.

## Clarity

<b>Title</b>	Clarity
<b>Long title</b>	Clarity of metadata
<b>Synonyms</b>	-
<b>Related</b>	Unambiguity, readability
<b>Definition</b>	The ease with which data consumers can understand the metadata.
<b>Indicators</b>	A grade (1-10)
<b>Examples on non-compliances</b>	<ul style="list-style-type: none"> <li>▪ The name of a file is 765897xyp.asc. This name has little meaning.</li> <li>▪ Data attribute 'profit' has no definition. It is not clear if it is net or gross profit.</li> </ul>
<b>Notes</b>	Other quality dimensions of metadata are completeness, correctness, and availability.

## Completeness (1)

<b>Title</b>	Completeness
<b>Long title</b>	Completeness of records
<b>Synonyms</b>	Coverage
<b>Related</b>	-
<b>Definition</b>	The degree to which all required records in the dataset are present.
<b>Indicators</b>	Percentage or number of the required records that are present.
<b>Examples on non-compliances</b>	<ul style="list-style-type: none"> <li>▪ Not all products are present in a product file.</li> <li>▪ Not all inhabitants of a city are recorded.</li> <li>▪ A file of trees also contains shrubs (over completeness)</li> </ul>
<b>Notes</b>	Incomplete records are also called missing units.

## Completeness (2)

<b>Title</b>	Completeness
<b>Long title</b>	Completeness of data values
<b>Synonyms</b>	-
<b>Related</b>	-
<b>Definition</b>	The degree to which all required data values are present.
<b>Indicators</b>	Percentage of the possible data values that are present.
<b>Examples on non-compliances</b>	<ul style="list-style-type: none"> <li>▪ In a product file the attribute supplier is not completed in every record.</li> <li>▪ In a questionnaire a respondent did not answer all questions.</li> </ul>
<b>Notes</b>	Incomplete data values are also called missing values.

## Consistency

<b>Title</b>	Consistency
<b>Long title</b>	Consistency of data values
<b>Synonyms</b>	-
<b>Related</b>	Plausibility of data values
<b>Definition</b>	The degree to which data values of two sets of attributes <ul style="list-style-type: none"><li>▪ within a record,</li><li>▪ within a data file,</li><li>▪ between data files,</li><li>▪ within a record at different points in time</li></ul> comply with a rule.
<b>Indicators</b>	Percentage of inconsistencies.
<b>Examples on non-compliances</b>	<ul style="list-style-type: none"><li>▪ A company is registered in the city of Paris in the country of Belgium.</li><li>▪ Overlap are gaps in file with address history of a person. For example: Address A from 1 Jan 2003 – 1 May 2019 and Address B from 1 March – until now.</li></ul>
<b>Notes</b>	-

## Currency

<b>Title</b>	Currency
<b>Long title</b>	Currency of data values
<b>Synonyms</b>	-
<b>Related</b>	Timeliness of availability of data
<b>Definition</b>	The degree to which data values are up to date.
<b>Indicators</b>	Percentage of data that are up to date in a point of time.
<b>Examples on non-compliances</b>	Outdated prices in the product file.
<b>Notes</b>	-

## Punctuality

<b>Title</b>	Punctuality
<b>Long title</b>	Punctuality of the availability of a dataset
<b>Synonyms</b>	-
<b>Related</b>	Timeliness of the availability of a dataset
<b>Definition</b>	The degree to which the period between the actual and target point in time of availability of a dataset is appropriate.
<b>Indicators</b>	<ul style="list-style-type: none"> <li>▪ The period between the actual and target point in time of availability of a dataset (days, hours, minutes).</li> <li>▪ Percentage of times that datasets were available too late (or too early).</li> </ul>
<b>Examples on non-compliances</b>	<ul style="list-style-type: none"> <li>▪ The dataset should be available at 1 July 2020 but is released at 3 July 2020. Too late.</li> <li>▪ The dataset should be available at 1 July 2020 at 10:00 am but is released at 9.45 am. Too early.</li> </ul>
<b>Notes</b>	<ul style="list-style-type: none"> <li>▪ A dataset can also exist of one transaction.</li> <li>▪ If no target time is agreed or planned, punctuality cannot be measured.</li> </ul>

## Timeliness

<b>Title</b>	Timeliness
<b>Long title</b>	Timeliness of the availability of a dataset
<b>Synonyms</b>	-
<b>Related</b>	Punctuality of the availability of a dataset
<b>Definition</b>	The degree to which the period between the time of creation of the real value and the time that the dataset is available is appropriate.
<b>Indicators</b>	Percentage of times a dataset was not available in a timely manner.
<b>Examples on non-compliances</b>	<ul style="list-style-type: none"> <li>▪ The date of birth of a person is after 23 days available in a dataset. It should be available within one week.</li> <li>▪ Data about quarterly returns are profit are available 3 month after the end of the quarter. The requirement is 1 month after the end of the quarter.</li> </ul>
<b>Notes</b>	<ul style="list-style-type: none"> <li>▪ Timeliness can only be measured if there is a norm for timeliness, e.g., one week after the event.</li> <li>▪ Timeliness is dependent on the duration of a process.</li> <li>▪ Data can be available punctual but not timely and the other way around.</li> </ul>

## Traceability

<b>Title</b>	Traceability
<b>Long title</b>	Traceability of data
<b>Synonyms</b>	-
<b>Related</b>	-
<b>Definition</b>	The degree to which data lineage is available.
<b>Indicators</b>	A grade (1-10)
<b>Examples on non-compliances</b>	The source of the data is unknown.
<b>Notes</b>	Data lineage is metadata that identifies the sources of data and the transformations through which it has passed up to the point of consumption.

## Uniqueness

<b>Title</b>	Uniqueness																		
<b>Long title</b>	Uniqueness of records																		
<b>Synonyms</b>	-																		
<b>Related</b>	Uniqueness of objects																		
<b>Definition</b>	The degree to which records occur only once in a data file.																		
<b>Indicators</b>	Percentage of duplicates in a data file.																		
<b>Examples on non-compliances</b>	Product A occurs twice in a file.																		
<b>Notes</b>	<ul style="list-style-type: none"> <li>▪ A record that occurs twice in a data file is called a duplicate.</li> <li>▪ Uniqueness of object is the degree to which objects (of the real world) occur only once as a record in a dataset.</li> </ul>																		
	<p>Three different problems can occur:</p> <p>a. One record with one key value occurs more than once in a dataset (duplicate with identical key values). The two records are not unique.</p> <table border="1" style="margin-left: 20px;"> <thead> <tr> <th>Key</th> <th>Name</th> </tr> </thead> <tbody> <tr> <td>22</td> <td>John</td> </tr> <tr> <td>22</td> <td>John</td> </tr> </tbody> </table> <p>b. One record with more than one key value occurs more than once in a dataset (duplicate with different key values). Object John is not unique in the dataset.</p> <table border="1" style="margin-left: 20px;"> <thead> <tr> <th>Key</th> <th>Name</th> </tr> </thead> <tbody> <tr> <td>22</td> <td>John</td> </tr> <tr> <td>37</td> <td>John</td> </tr> </tbody> </table> <p>c. One record has the same key as another record, and both occur in a dataset (false duplicate). Key 22 is not unique.</p> <table border="1" style="margin-left: 20px;"> <thead> <tr> <th>Key</th> <th>Name</th> </tr> </thead> <tbody> <tr> <td>22</td> <td>John</td> </tr> <tr> <td>22</td> <td>Peter</td> </tr> </tbody> </table>	Key	Name	22	John	22	John	Key	Name	22	John	37	John	Key	Name	22	John	22	Peter
Key	Name																		
22	John																		
22	John																		
Key	Name																		
22	John																		
37	John																		
Key	Name																		
22	John																		
22	Peter																		

## Validity

<b>Title</b>	Validity
<b>Long title</b>	Validity of data values
<b>Synonyms</b>	-
<b>Related</b>	Accuracy of data values Completeness of data values Consistency of data values
<b>Definition</b>	The degree to which data values comply with rules.
<b>Indicators</b>	Percentage of data values that do not comply with rules.
<b>Examples on non-compliances</b>	<ul style="list-style-type: none"><li>▪ A city that does not exist in a list of cities.</li><li>▪ A birth data that is out of range of valid birth dates.</li></ul>
<b>Notes</b>	<ul style="list-style-type: none"><li>▪ A data value can be valid but not accurate.</li><li>▪ A data value can be valid but incomplete. I can be allowed that a data value is not known.</li><li>▪ A valid data value is part of a value domain.</li><li>▪ Consistency is about comparing two or more data values.</li></ul>

## Appendix 4: Definitions of concepts and data concepts

In this Appendix concepts are defined that are relevant in this report. See Table 6.

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

Each word that appears in **bold** in the definition of a concept is a concept defined elsewhere in Table 6. This way the coherence between the concepts are made visible.

**Table 6:** Definitions of concepts

Concept	Definition	Source	Relationships with other concepts
Attribute	A <b>characteristic</b> of an <b>entity type</b> about which the organisation wishes to hold information.	-	Distinguishes <b>entity type</b> Is specified by its name, definition, classification and <b>format</b> .
Characteristic	Distinguishing feature	ISO 9000	-
Composition of a dataset	The way in which a <b>dataset</b> is made up.	-	
Concept	Unit of knowledge created by a unique combination of <b>characteristics</b>	ISO 1087	-
Concept system	A set of <b>concepts</b> 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 <b>data</b> according to the purpose for which it is used.	-	-
Data concept	A <b>concept</b> related to <b>data</b>	-	Has associated <b>dimensions</b>
Data file	<b>Data</b> stored on a computer as one unit with one name.	Cambridge 2020	Is part of a <b>dataset</b> .
Data item	One occurrence of an <b>attribute</b>	-	Contains <b>data value</b>
Data lineage	<b>Metadata</b> that identifies the sources of <b>data</b> and the transformations through which it has passed up to the point of consumption.	-	-
Data pattern	A series of <b>data</b> that repeats in a recognizable way.	Investopedia	-

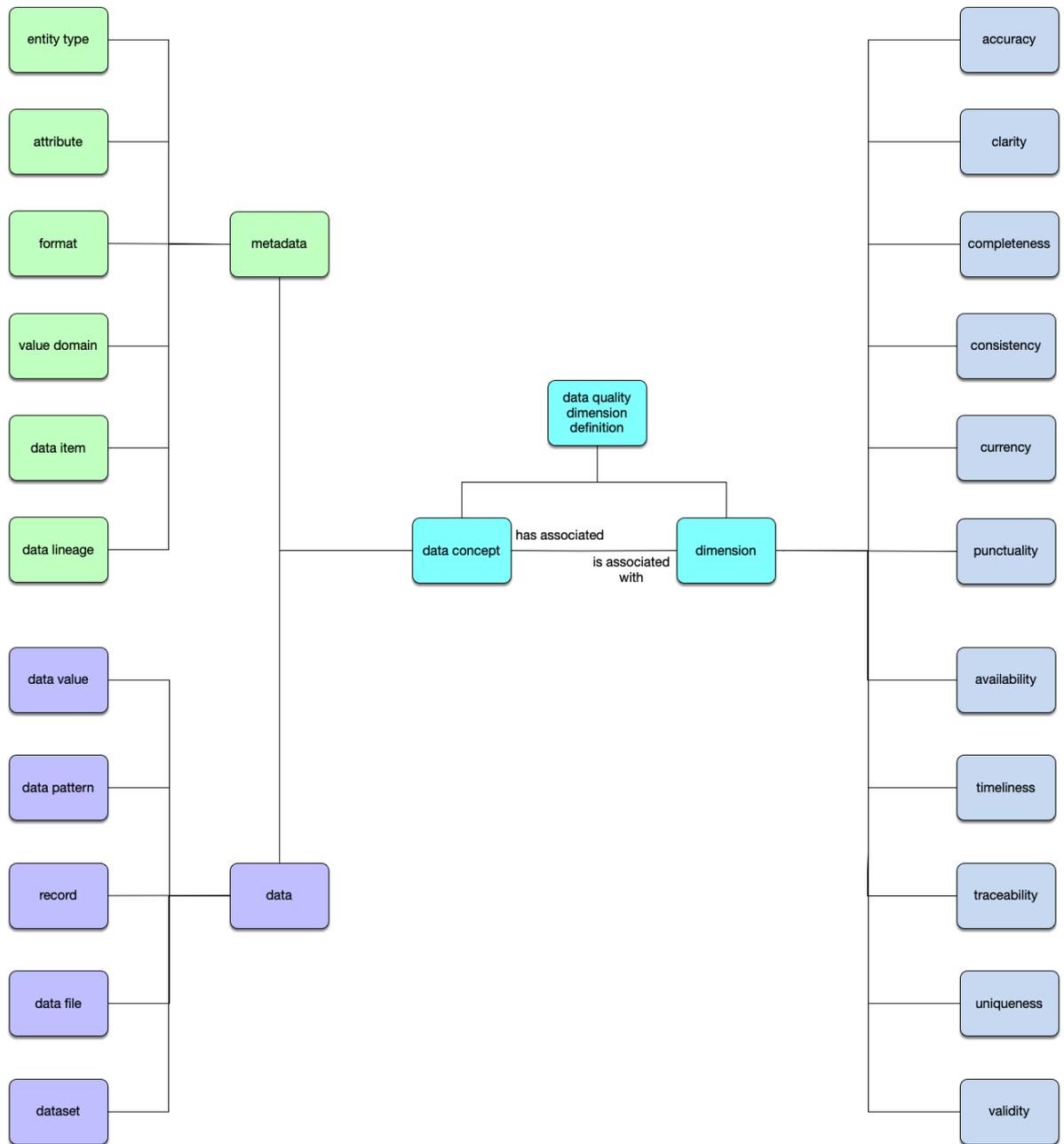
Concept	Definition	Source	Relationships with other concepts
Data value	The value of a <b>data item</b> .	-	Is contained in <b>data item</b> Forms part of <b>record</b> Is within <b>value domain</b> Represents a <b>property</b> of an <b>object</b>
Dataset	Any organized collection of <b>data</b> .	Early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

Concept	Definition	Source	Relationships with other concepts
Register	A <b>dataset</b> 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 <b>data category</b> .
Statistical output	Output from a statistical process.	-	Is an instance of <b>data category</b> .
Transactional data	Data that describes an event that takes place as an organization conducts its business.	-	Is an instance of <b>data category</b> .
Real value	The real-life value of a <b>property</b> of an <b>object</b> .	-	Expresses an instance of a <b>property</b> .
Reference data	<b>Data</b> used to categorize other data.	-	Is an instance of <b>data category</b> .
Record	A logically related set of <b>data values</b> that represent a (real-world) <b>object</b>	-	Forms part of <b>data file</b> Is composed of <b>data values</b>
Value domain	A set of permissible values of an <b>attribute</b> .	-	Includes <b>data value</b>

Source: Black, A., Nederpelt, P. van. (2020). *Data concept system for Data Quality Dimensions*. Research paper. DAMA-NL.

## Appendix 5: Diagrams

Figure 2 shows that a dimension is associated with a data concept. The definition of a dimension of data quality is formed by the combination of a dimension and a data concept. In the diagram, only the common dimensions are presented.



**Figure 2:** Relationship between data concepts and dimensions

Figure 3 is an artist's impression of the real world and data world.

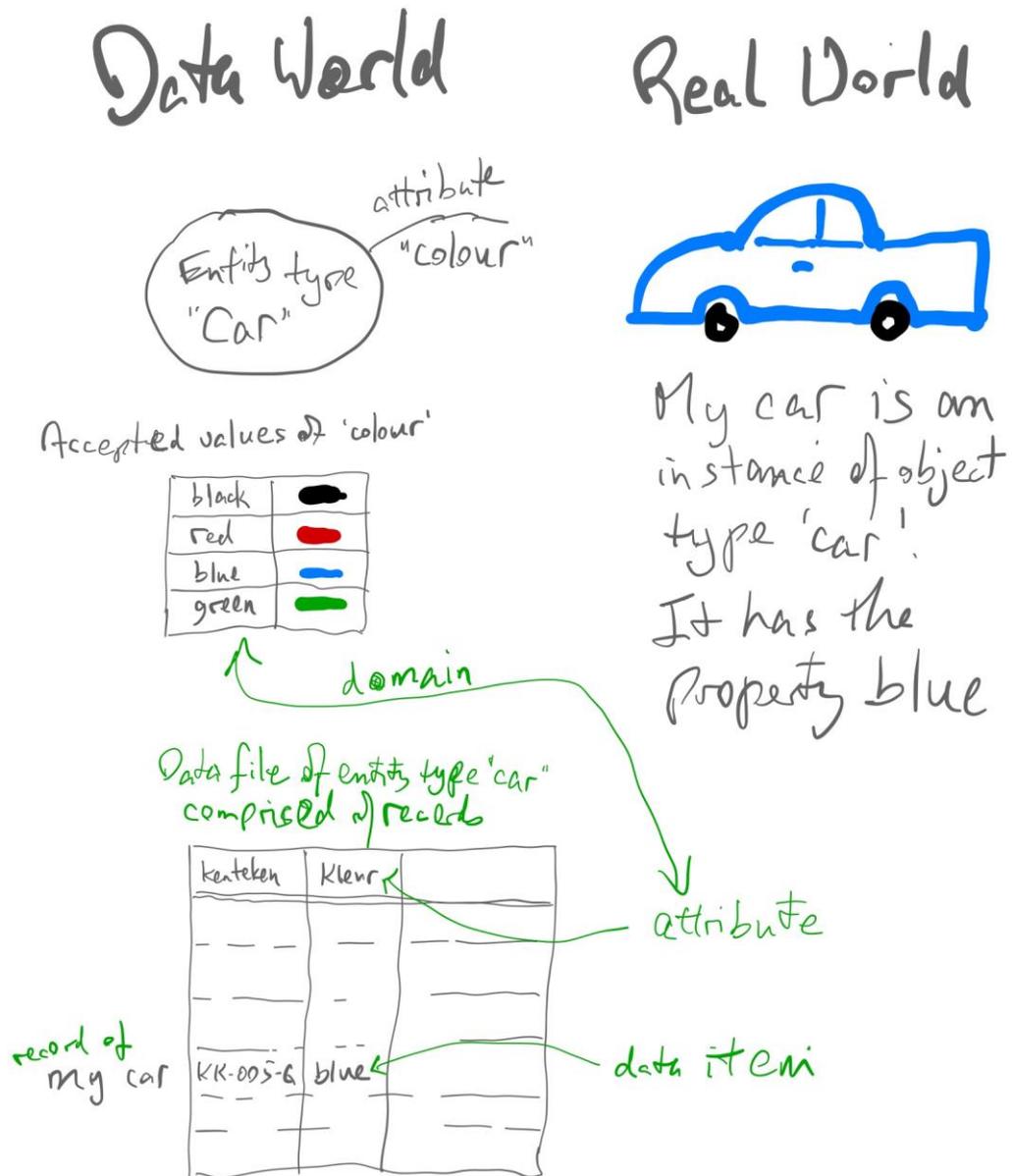


Figure 3: Artist's impression of the real world and data world

## Appendix 6: Sources

### Sources of definitions of quality dimension

- Brackett, M. H. (2012). *Data Resource Integration: Understanding and Resolving a Disparate Data Resource* (1<sup>st</sup> 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](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*. 2<sup>nd</sup> 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* (2<sup>nd</sup>ed.). 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.

### Other sources

- Black, A., Nederpelt, P. van. (2020). *Data concept system for Data Quality Dimensions*. *Research paper*. DAMA-NL.
- Black, A., Nederpelt, P. van. (2020). *Dimensions of Data Quality Dimensions*. *Research paper*. DAMA-NL.
- Brackett, M. H. (2012). *Data Resource Integration: Understanding and Resolving a Disparate Data Resource* (1<sup>st</sup> 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*. 2<sup>nd</sup> Edition. Technics Publications Llc. August 2017.
- Earley, S. (2011). *The DAMA Dictionary of Data Management* (2<sup>nd</sup> ed.). NJ: Technics Publications Llc.

- 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>
- 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](https://www.unece.org/trade/untdid/texts/d300_s.htm)

## Version history

Version	Date	Description of the modification	Author
1.0.p1	14 August 2020	First draft	Peter
1.0.p2	20 August	Amendments and comments	Andrew
1.0.p3	20 August	Amendments and comments processed	Peter
1.0.p4	27 August 2020	Comments Fred Dijk processed.	Peter
1.0.p5	28 Augustus 2020	Diagram edited.	Andrew
1.0.p6	28 August 20	Amendments and comments	Andrew
1.0	3 Sept 20	Comments processed	Peter

Active distribution per version	
Version	Distribution
1.0.p1-2	Dropbox
1.0.p3	Dropbox, Fred Dijk (review)
1.0.p4	Dropbox, Reviewers.
1.0.p5-p6	Dropbox
1.0	Dropbox. Website DAMA-NL