

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)

Reviewers:

Fred Dijk (Scamander)

Aris Prins (Premium)

Version: See Version History

Printed of 15-11-2020 14:41

© DAMA NL Foundation. All rights reserved.

| | |
|---|-----------|
| Table of contents | |
| Tables | 4 |
| Figures | 4 |
| 1. Introduction | 5 |
| 1.1 The importance of data quality | 5 |
| 1.2 Purpose of this document | 5 |
| 1.3 Audience of this document | 5 |
| 1.4 How did the document come about? | 5 |
| 1.5 Release policy | 5 |
| 1.6 Reading guideline | 5 |
| 2. Definitions and examples | 6 |
| 2.1 What is data quality? | 6 |
| 2.2 What is a data concept? | 6 |
| 2.3 What is a dimension? | 7 |
| 2.4 Combination of a dimension and a data concept | 8 |
| 2.5 What is a requirement? | 8 |
| 3. How to select the right dimensions of data quality? | 9 |
| 3.1 Step 1: Determine which dimensions of data quality are important for the data under consideration | 9 |
| 3.2 Step 2: Determine whether a dimension contributes sufficiently to a higher objective | 9 |
| 3.3 Step 3: Prioritize the dimensions | 9 |
| 3.4 Step 4: Establish indicators and associated measurement methods for the selected dimensions | 9 |
| Appendix 1: Dimensions of data quality | 10 |
| Appendix 2: Combinations of dimensions and data categories | 17 |
| Appendix 3: Elaborated dimensions of data quality | 20 |
| 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 |
| Appendix 4: Definitions of concepts and data concepts | 27 |
| Appendix 5: Diagrams | 30 |
| Appendix 6: Sources | 32 |
| Sources of definitions of quality dimension | 32 |
| Other sources | 32 |
| Version history | 34 |

Tables

| | |
|--|----|
| Table 1: Examples of combinations of dimensions and data concepts..... | 8 |
| Table 2: Examples of requirements of data quality | 8 |
| Table 3: Definitions of dimensions of data quality..... | 11 |
| Table 4: Combinations of dimensions and data categories..... | 17 |
| Table 5: Definitions of concepts | 27 |

Figures

| | |
|--|----|
| Figure 1: Data concepts in a data model..... | 7 |
| Figure 2: Relationship between data concepts and dimensions | 30 |
| Figure 3: Artists impression of the real world and data world..... | 31 |

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.

Bad quality data puts an organisation at risk. It can lead to bad decisions, unsatisfied customers, unsatisfied data consumers, fines due to non-compliance, hidden costs (rework), bad reputation, unsatisfied employees, and lack of interoperability.

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.

It offers, therefore, a list of 60 dimensions of data quality and corresponding, standardized definitions (Appendix 1). The relationship with existing definitions can be found in our exhaustive research report Dimensions of Data Quality (Black & Van Nederpelt, 2020).

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 version of the report was published in September 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.

1.6 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.

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?

ISO 9000:2015 defines quality as:

Quality is the degree to which inherent characteristics of an object meet requirements.

We derive from this definition the following definition of data quality:

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

Note 1: The term characteristics in the definition of ISO 9000 is substituted by dimensions, because this term is more common in data management.

Note 2: The adjective inherent is left out, because also extrinsic dimensions are relevant in data management such as availability.

Note 2: Data take various forms: data concepts. Each dimension of data quality is defined in relation to such a data concept.

2.2 What is a data concept?

A data concept is defined as:

Data concept is a form by which data is structured and organised in an information system.

Some examples of data concepts are dataset, data file, record, attribute, and data value.

Data concepts and their definitions can be found in Appendix 4.

A subset of data concepts is shown in a data model in Figure 1.

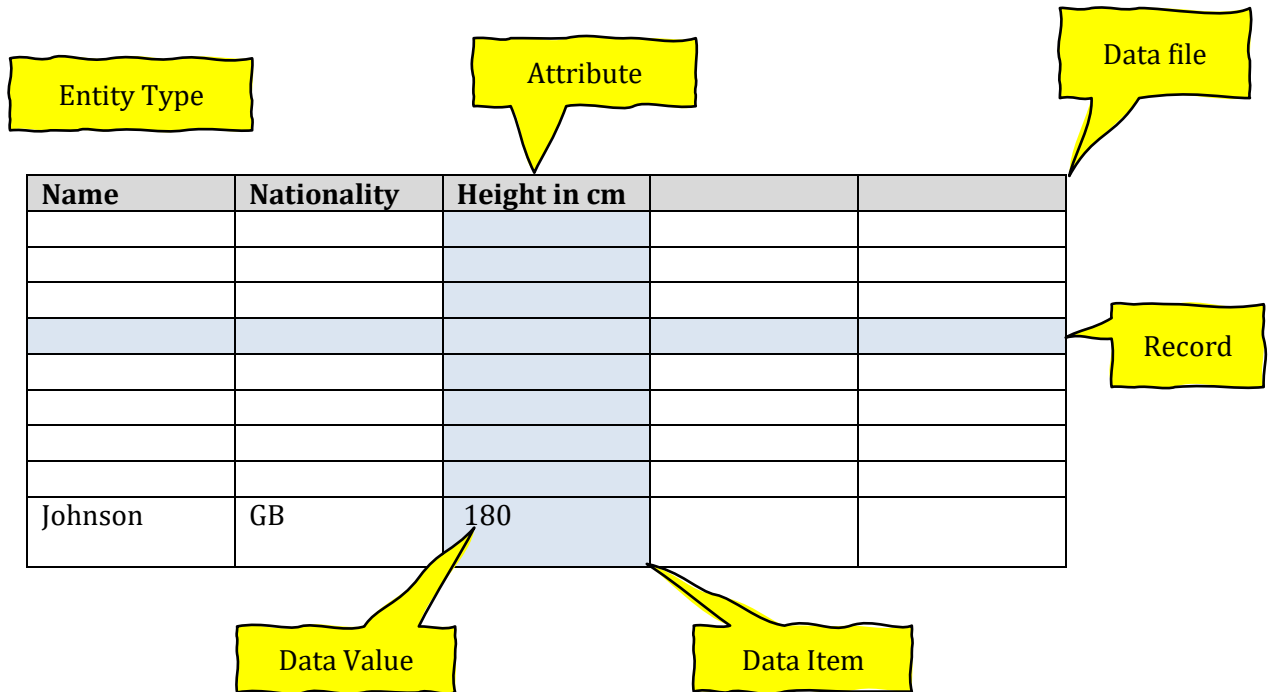


Figure 1: Data concepts in a data model

In Appendix 5, two other diagrams of data concepts are presented.

2.3 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.

2.4 Combination of a dimension and a data concept

A dimension and a data concept should be a logical combination. Examples of these combinations are presented in Table 1. An illogical combination is, e.g., accuracy of a data file.

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) |

Definitions of dimensions of data quality in this report are composed of a combination of a dimension and a data concept. In everyday language we tend to only mention the dimension without the associated data concept. See also Figure 2 in Appendix 5.

2.5 What is a requirement?

Requirement is a need or expectation that is stated, generally implied or obligatory (ISO 9000).

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

| Dimension | Data concept | Requirement |
|-----------------------|--------------|---|
| 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 their stakeholders. 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 the first steps in a procedure to control or improve data quality. These steps are:

- Determine which dimensions of data quality are important for the data under consideration.
- Determine whether a dimension contributes sufficiently to a higher objective.
- Prioritize the selected dimensions
- Establish indicators and associated measurements methods for the selected dimensions

These steps are elaborated below.

The persons who have interests in the data should be involved in the selection process. Namely those who are responsible, accountable, consulted, and informed (RACI) about data quality according to their roles.

3.1 Step 1: Determine which dimensions of data quality are important for the data under consideration

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

Determine which dimensions are important for the data category. Table 4 in Appendix 2 indicates which dimensions are candidates for a specific data category.

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

Determine whether a dimension contributes sufficiently to a higher objective, i.e. to some business goal. The contribution must be large enough to make it worthwhile to select the dimension.

The following are examples of such 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, regulations, and requirements of regulators.
- I. Data driven decisions making

See Appendix 1 for 60 dimensions of data quality and their definitions.

3.3 Step 3: Prioritize the dimensions

Rank the dimensions in order of priority. Put the dimension with the best cost-benefit ratio on the first place, thereby keeping control over your costs.

3.4 Step 4: Establish indicators and associated measurement methods for the selected dimensions

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

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 3 shows the dimensions, the associated data concept, and their definitions in alphabetical order.

The last column presents the unit of measure:

| Unit of measure | Remark |
|-----------------|---|
| % | Percentage |
| Number | Absolute number |
| Grade | Only the perception of people about the dimension can be measured. A grade can be number on a scale of 0-10 or 1-5. |
| Boolean | Yes/no or true/false |
| Duration | Expressed in seconds, minutes, hours, days, weeks or months. |
| Story | The value of the dimension cannot be expressed in a number and should be explained in a 'story'. |

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 3: Definitions of dimensions of data quality

| Nr | Dimension | Data Concept | Definition | Unit of measure |
|-----------|----------------------------------|-------------------------|---|------------------------|
| 1. | Access security | Datasets | The degree to which access to datasets is restricted. | Grade |
| 2. | Accessibility | Data | The ease with which data can be consulted or retrieved, | Grade |
| 3. | Accuracy | Data values | The degree of closeness of data values to real values. | % |
| 4. | Appropriateness | Format | The degree to which the format is suitable for use. | % |
| 5. | Availability | Data | The degree to which data can be consulted or retrieved by data consumers or a process. | Grade |
| 6. | Ability to represent null values | Format | The degree to which a format allows null values in an attribute. | Yes/No |
| 7. | Clarity | Metadata | The ease with which data consumers can understand the metadata. | Grade |
| 8. | Coherence | Composition of datasets | The degree to which datasets can be combined. | Story |
| 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. | Grade |
| 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. | Grade |
| 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. | % |

| Nr | Dimension | Data Concept | Definition | Unit of measure |
|-----|---|--|---|-----------------|
| 13. | | Data files | The degree to which all required data files are present. | %, Number |
| 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. | Story |
| 18. | | Composition of datasets | The degree to which the composition of datasets is in accordance with laws, regulations, or standards. | Story |
| 19. | Confidentiality | Data | The degree to which disclosure of data should be restricted to authorized data consumers. | Grade |
| 20. | Consistency | Data values | The degree to which data values of two sets of attributes <ul style="list-style-type: none"> ▪ within a record, ▪ within a data file, ▪ between data files, ▪ within a record at different points in time 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. | % |

| Nr | Dimension | Data Concept | Definition | Unit of measure |
|-----------|------------------|--|---|------------------------|
| 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. | % |
| 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. | Grade |
| 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. | Story |
| 29. | | Records | The degree to which objects are aggregated to records. | Story |
| 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. | Duration |
| 33. | Linkability | Data files | The degree to which records of one data file can be correctly coupled with records of another data file. | % |

| Nr | Dimension | Data Concept | Definition | Unit of measure |
|-----------|---------------------|-------------------------|---|-----------------------------|
| 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. | Grade |
| 36. | Objectivity | Data values | The degree to which the data values are created in an unbiased manner. | Grade |
| 37. | Obtainability | Data | The degree to which the data can be acquired. | Grade |
| 38. | Plausibility | Data values | The degree to which data values match knowledge of the real world. | Story |
| 39. | Portability | Data | The degree to which data can be installed, replaced, or moved from one system to another while preserving the existing quality. | Story |
| 40. | Portability | Format | The degree to which a format can be applied in a wide range of situations. | Story |
| 41. | Precision (1) | Data values | The degree of accuracy with which data values are recorded or classified. | Depends on data or metadata |
| 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. | Duration |
| 44. | Reasonability | Data pattern | The degree to which a data pattern meets expectations. | Grade |

| Nr | Dimension | Data Concept | Definition | Unit of measure |
|-----------|-----------------------|-------------------------|--|------------------------|
| 45. | Recoverability | Datasets | The degree to which datasets are preserved in the event of incident. | Story |
| 46. | Redundancy | Data | The degree to which logically identical data are stored more than once. | Number |
| 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. | Story |
| 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. | Story |
| 51. | Reputation | Data | The degree to which data are trusted or highly regarded in terms of their source or content. | Grade |
| 52. | Retention period | Datasets | The period that datasets are available until they can or must be deleted. | Duration |
| 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. | Duration |
| 54. | Traceability | Data | The degree to which data lineage is available. | Story |
| 55. | Uniqueness | Objects | The degree to which objects (of the real world) occur only once as a record in a data file. | % |

| Nr | Dimension | Data Concept | Definition | Unit of measure |
|-----------|------------------|---------------------|---|------------------------|
| 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. | Grade |
| 59. | Variety | Data | The degree to which data are available from different data sources. | Story |
| 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 4 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.

Column A contains the selection of DAMA-UK as expressed in the Six Primary Dimensions for Data Quality Assessment.

Column B contains the selection as expressed in the List of Conformed Dimensions of Data Quality.

Table 4: Combinations of dimensions and data categories

| Nr | Dimension | Data Concept | A | B | Data Category | | | | |
|----|----------------------------------|-----------------------------|---------|----------------|---------------|----------------|--------------------|-----------|--------------------|
| | | | DAMA UK | CDDQ | Master Data | Reference Data | Transactional Data | Registers | Statistical Output |
| 1 | Access security | Datasets | | | X | X | X | X | |
| 2 | Accessibility | Data | X | | | | | | 9 |
| 3 | Accuracy | Data values | X | X | X | X | X | X | 2 |
| 4 | Appropriateness | Format | | | | | | | |
| 5 | Availability | Data | | | | | | | |
| 6 | Ability to represent null values | Format | | | | | | | |
| 7 | Clarity | Metadata | | X ¹ | 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 | X | |
| 13 | Completeness | Data files | X | | | | | | |
| 15 | Completeness | Data values of an attribute | X | | | | | | |
| 12 | Completeness | Records | X | | X | X | X | X | |
| 11 | Completeness | Attributes | X | | | | | | |
| 16 | Completeness | Metadata | | | X | X | X | X | |

¹ In CDDQ clarity is called representation.

| Nr | Dimension | Data Concept | A | B | Data Category | | | | |
|----|---|--|---------|------|---------------|----------------|--------------------|-----------|--------------------|
| | | | DAMA UK | CDDQ | Master Data | Reference Data | Transactional Data | Registers | Statistical Output |
| 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 | X | X | | | | | |
| 21 | Consistency | Data values of a set of attributes of a dataset at different points in time (temporal consistency) | X | X | | | | | |
| 22 | Consistency | Data values of two sets of attributes between datasets (across datasets) | X | X | | | | | |
| 23 | Consistency | Data values of two sets of attributes between records (cross record) | X | X | | | | | |
| 23 | Consistency | Data values of two sets of attributes within a record (record level) | X | X | | | | | |
| 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 | X | | | | | | |
| 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 | | X | | | | | |

| Nr | Dimension | Data Concept | A | B | Data Category | | | | | |
|----|-----------------------|-------------------------|---------|----------------|---------------|----------------|--------------------|-----------|--------------------|---|
| | | | DAMA UK | CDDQ | Master Data | Reference Data | Transactional Data | Registers | Statistical Output | |
| 42 | Precision (2) | Data values | | | | | | | | |
| 43 | Punctuality | Dataset availability | | | | | | | | 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 |
| 49 | Reliability | Initial data value | | | | | | | | 3 |
| 50 | Reproducibility | Dataset | | | | | | | | |
| 51 | Reputation | Data | | | | | | | | |
| 52 | Retention period | Datasets | | | | | X | | | |
| 53 | Timeliness | Dataset availability | X | X | | | | | | 4 |
| 54 | Traceability | Data | | X ² | | | X | | | |
| 56 | Uniqueness | Records | X | | X | X | | X | | |
| 55 | Uniqueness | Objects | X | | | | | X | | |
| 57 | Validity | Data values | X | X | | | X | X | | |
| 58 | Value | Data | | | | | | | | |
| 59 | Variety | Data | | | | | | | | |
| 60 | Volatility | Data values | | | | | | | | |

² In CDDQ traceability is called data lineage.

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-compliance with required data quality.
- Notes

Accuracy

| | |
|-----------------------------------|---|
| Title | Accuracy |
| Long title | Accuracy of data values |
| Synonym | Correctness of data values |
| Related | - |
| Definition | The degree of closeness of data values to real values. |
| Indicators | Percentage or number of inaccurate data values. |
| Examples of non-compliance | <ul style="list-style-type: none"> ▪ A house is located at number 120 but registered as number 12. ▪ A person is called Janssen but registered as Jansen. ▪ 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. ▪ A product is located at A23 but according to the database its location is P76. ▪ 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. |
| Notes | <ul style="list-style-type: none"> ▪ The data producer or consumer must define when he/she considers a data value as inaccurate and define criteria for inaccuracy. ▪ The impact of an inaccuracy is different for each attribute. ▪ Generally, accuracy will be measured for individual attributes, e.g., the accuracy of the product name. |

Availability

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

Clarity

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

Completeness (1)

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

Completeness (2)

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

Consistency

| | |
|-----------------------------------|--|
| Title | Consistency |
| Long title | Consistency of data values |
| Synonyms | - |
| Related | Plausibility of data values |
| Definition | The degree to which data values of two sets of attributes <ul style="list-style-type: none">▪ within a record,▪ within a data file,▪ between data files,▪ within a record at different points in time comply with a rule. |
| Indicators | Percentage of inconsistencies. |
| Examples of non-compliance | <ul style="list-style-type: none">▪ A company is registered in the city of Paris in the country of Belgium.▪ 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. |
| Notes | - |

Currency

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

Punctuality

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

Timeliness

| | |
|-----------------------------------|---|
| Title | Timeliness |
| Long title | Timeliness of the availability of a dataset |
| Synonyms | - |
| Related | Punctuality of the availability of a dataset |
| Definition | 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. |
| Indicators | Percentage of times a dataset was not available in a timely manner. |
| Examples of non-compliance | <ul style="list-style-type: none"> ▪ The date of birth of a person is available in a dataset after 23 days. It should be available within one week. ▪ Data about quarterly returns of VAT are available 3 months after the end of the quarter. The requirement is 1 month after the end of the quarter. |
| Notes | <ul style="list-style-type: none"> ▪ Timeliness can only be measured if there is a norm for timeliness, e.g., one week after the event. ▪ Timeliness is dependent on the duration of a process. ▪ Data can be available punctually but not timely and the other way around. |

Traceability

| | |
|-----------------------------------|--|
| Title | Traceability |
| Long title | Traceability of data |
| Synonyms | - |
| Related | - |
| Definition | The degree to which data lineage is available. |
| Indicators | A grade (1-10) |
| Examples of non-compliance | The source of the data is unknown. |
| Notes | 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

| Title | Uniqueness | | | | | | | | | | | | | | | | | | |
|-----------------------------------|--|-----|------|----|------|----|------|-----|------|----|------|----|------|-----|------|----|------|----|-------|
| Long title | Uniqueness of records | | | | | | | | | | | | | | | | | | |
| Synonyms | - | | | | | | | | | | | | | | | | | | |
| Related | Uniqueness of objects | | | | | | | | | | | | | | | | | | |
| Definition | The degree to which records occur only once in a data file. | | | | | | | | | | | | | | | | | | |
| Indicators | Percentage of duplicates in a data file. | | | | | | | | | | | | | | | | | | |
| Examples of non-compliance | Product A occurs twice in a file. | | | | | | | | | | | | | | | | | | |
| Notes | <ul style="list-style-type: none"> ▪ A record that occurs twice in a data file is called a duplicate. ▪ Uniqueness of object is the degree to which objects (of the real world) occur only once as a record in a dataset. | | | | | | | | | | | | | | | | | | |
| | <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

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

Appendix 4: Definitions of concepts and data concepts

In this Appendix concepts are defined that are relevant in this report. See Table 5.

A distinction is made between data concepts in the real world (purple) and the 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 5. This way the coherence between the concepts is made visible.

Table 5: Definitions of concepts

| 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 form by which data is structured and organised in an information system. | - | 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 | - |

| Concept | Definition | Source | Relationships with other concepts |
|----------------------|---|-----------------------|--|
| Data quality | Data quality is the degree to which dimensions of data meet requirements | Adapted from ISO 9001 | - |
| 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 | - |
| 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 |

| Concept | Definition | Source | Relationships with other concepts |
|--------------------|---|----------|--|
| Property | A feature of an object . | ISO 1087 | Characterises object Is recorded by data value Actually, has real value |
| 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 |

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.

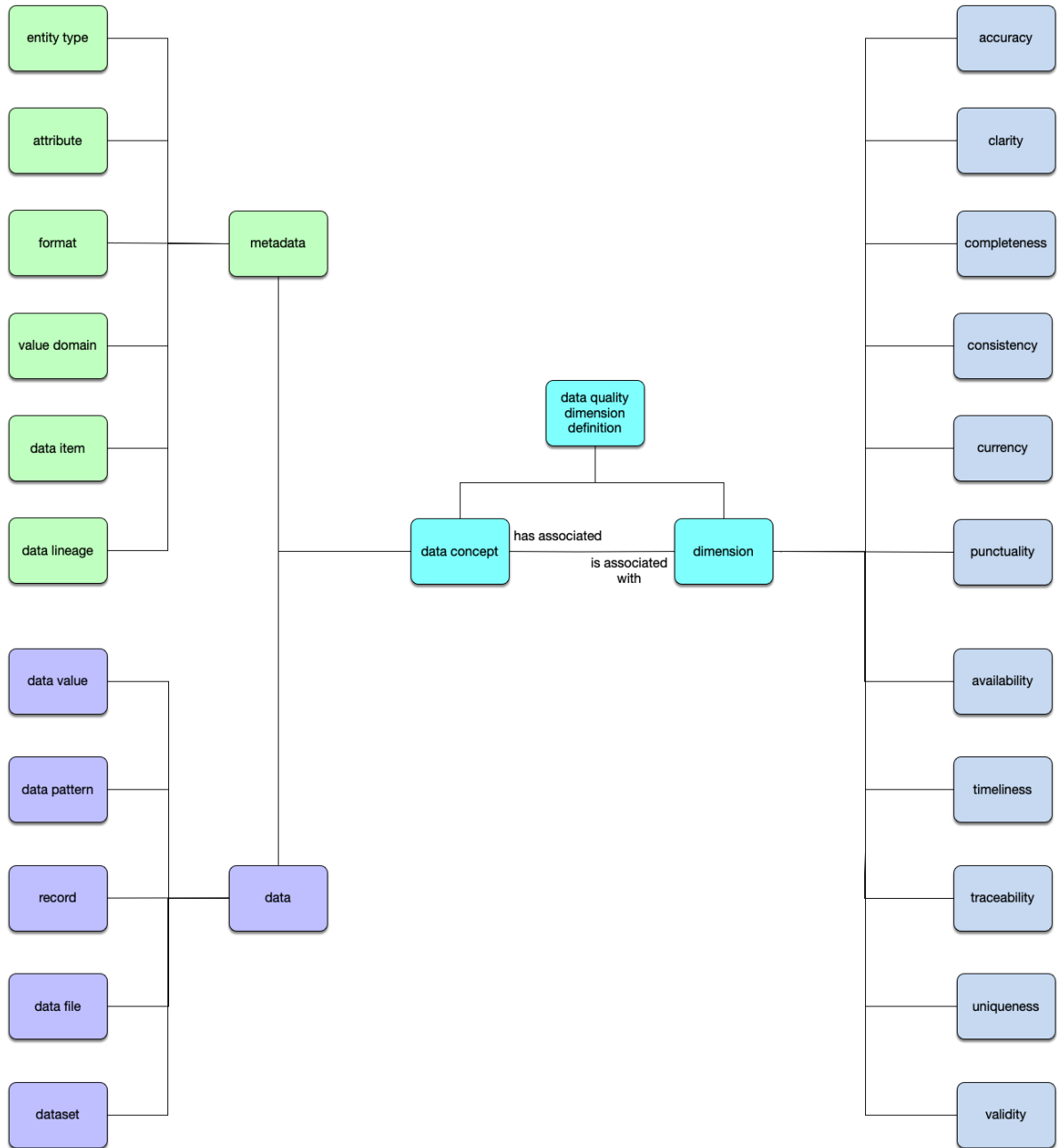


Figure 2: Relationship between data concepts and dimensions

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

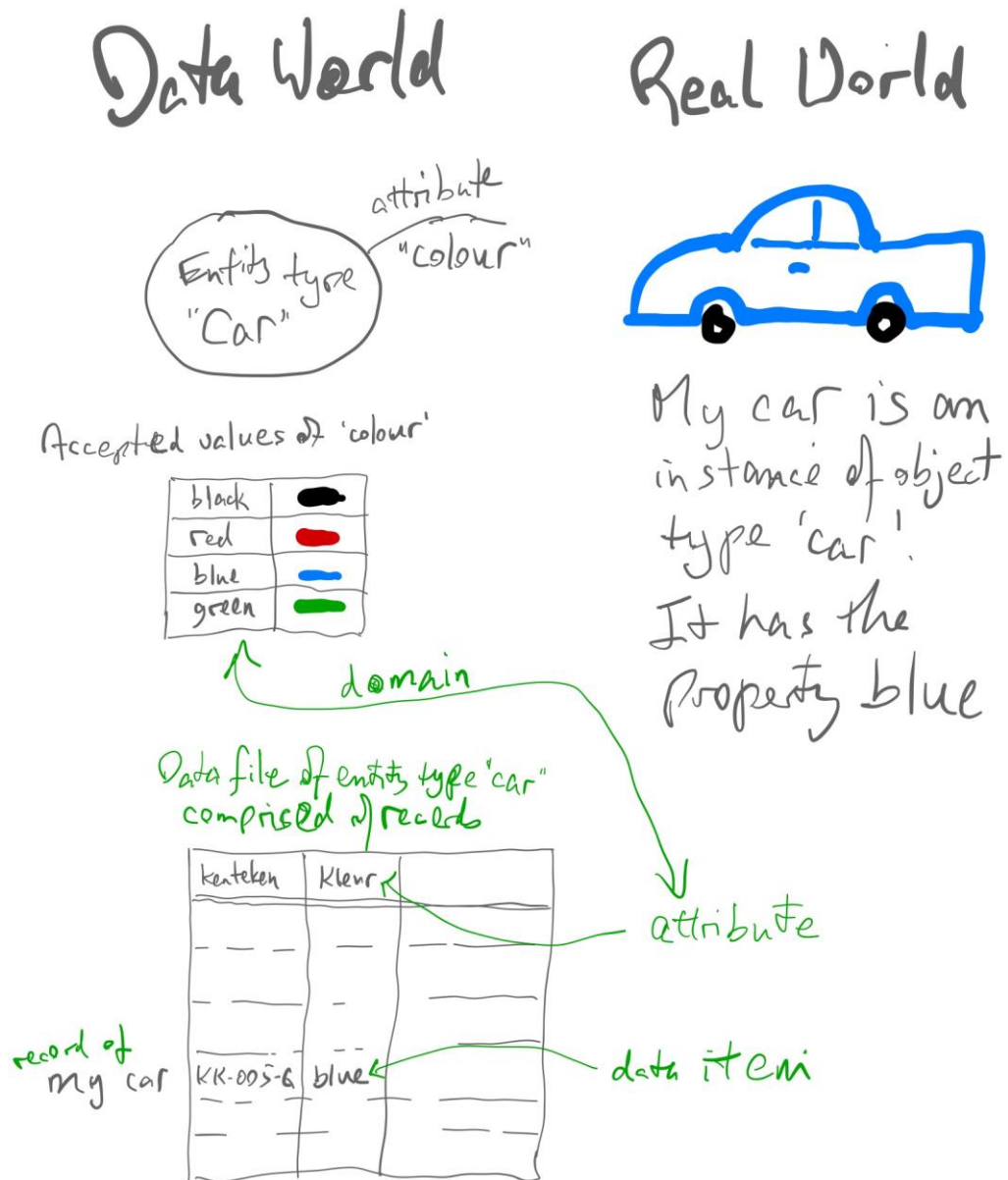


Figure 3: Artists impression of the real world and data world

Appendix 6: Sources

Sources of definitions of quality dimension

- Brackett, M. H. (2012). *Data Resource Integration: Understanding and Resolving a Disparate Data Resource* (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.

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* (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.

- 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

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 |
| 1.1.p1 | 19 Sept 20 | Added: <ul style="list-style-type: none"> • Risks of insufficient data quality • Reference to research report DDQ with existing definitions • Unit of measure • Roles and responsibilities (RACI) • Excel spreadsheet • Names of reviewers • Prioritization of dimensions Removed: <ul style="list-style-type: none"> • Procedure to improve data quality. | Peter |
| 1.1.p1 | 20 Sept 20 | Amendments and comments | Andrew |
| 1.1.p3 | 21 Sept 20 | Amendments and comments processed. Chapter 2 reformulated. | Peter |
| 1.1.p4 | 1 Oct 20 | Amendments to definitions | Andrew |
| 1.1 | 14 Nov 20 | Amendments 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 |
| 1.1.p1 | Dropbox. Fred Dijk. |
| 1.1.p2 | Dropbox |
| 1.1.p3 | Dropbox, Fred Dijk |
| 1.1.p4 | Dropbox |
| 1.1 | Dropbox. Website DAMA-NL |