Dimensions of Data Quality (DDQ)

Research Paper

Table of contents

Tables

Table 1: Definitions of dimensions of data quality
Table 2: Definitions of dimensions of data quality classified by data concept
Table 3: Definitions of concepts
Table 4: Criteria for definitions
Table 5: Dimensions of data quality, classification, and sources

Figures

Figure 1: Data concept system DDQ
Figure 2: Data concepts in a data model
Figure 3: Relationship between data concepts and dimensions
Figure 4: Artists impression of the real world and data world

Foreword

Recorded data has been important ever since people started trading. The earliest known writing was a record of quantities of livestock and commodities. Systems such as double-entry bookkeeping were developed to ensure the accuracy of critical data. Early forms of this were pioneered by the Romans and in the Jewish community of the early-medieval Middle East. The problem of ensuring acceptable data quality escalated with the rise of automated systems that record and process growing volumes of data. Early efforts to actively improve data quality confronted the problem of definition. How do we specify data quality and how do we measure it? Basic measurements are normally done on dimensions. So, the concept of 'data quality dimensions' was born.

In the last thirty years many different sets of data quality dimensions have been described with diverse definitions for each dimension. In this publication the authors have assembled a comprehensive survey of published dimensions of data quality. From these standardized definitions have been formulated and a selection of preferred dimensions has been made.

DAMA NL
5 August 2020

About the authors

Andrew Black

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

Peter van Nederpelt

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

1. Introduction

1.1 Importance

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.

However, there is still no standard for the dimensions of data quality and their definitions. The purpose of this report is to take a step towards a standard. Such a standard does exist in the domain of statistical institutions; although it has not been published as a separate product.

1.2 Dimensions, data and quality

The title of this report is Dimensions of Data Quality (DDQ).

Dimension is defined as a measurable feature of an object (ISO 9001). The object is in this context data. The term dimension is used to make the connection to dimensions in the measurement of physical objects (e.g., length, width, height).

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.

Data is a collective term for several concepts such as records, attributes, data values and metadata.

Quality is defined as the degree to which inherent characteristics of an object meet requirements (ISO 9001). But we reformulate this definition as: the degree to which dimensions of a data concept meet requirements.

The term 'characteristic' is replaced by 'dimension', because the latter term is more common in data land.

The term 'object' is replaced by 'data concept' to avoid confusion with objects of the real world that are represented in datasets. The term concept is, moreover, used in ISO 704, a standard for definitions.

1.3 Scope

The scope of the report is the quality dimensions of data concepts that occur in information systems.

1.4 Research question

The research question is:
What definitions of dimensions of data quality are to be preferred?

Sub questions are:
- What dimensions of data quality can be distinguished?
- What definitions of data quality dimension have been published?
- Which definitions are to be preferred?
- Which set of dimensions are essential?
- How can dimensions be classified?

1.5 Target group

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

1.6 Background, ownership, and management

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

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

1.7 Release policy

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

1.8 Reading guide

Chapter 1 describes the scope of the report, the research question, the intended audience, the motivation for writing the report and the management of the report after publication.
Chapter 2 describes the methodology: the steps that have been taken to arrive at a list of preferred definitions.
Chapter 3 presents the final result of the study.
Chapter 4 concludes whether the research question has been answered and considers what can be done with the result.
Appendix 1 contains references to sources for definitions.
Appendix 2 defines concepts used in the report
Appendix 3 shows diagrams of the concept system.
Appendix 4 contains criteria for definitions of quality dimensions.
Appendix 5 is a list of quality dimensions and their sources.
Appendix 6 provides an overview of definitions per dimension. The preferred definitions are also selected here.

2. Methodology

This chapter describes the input used in the analysis and the analysis steps taken.

2.1 Step 1: Collect sources

- Draw up a list of data quality dimensions and their source.
- These are in any case the sources of definitions of dimensions mentioned in DMBOK (2017) Chapter 13 on Data Quality.
- Use the report Concept System of Dimensions of Data Quality (Black, Van Nederpelt, 2020).
- Use a standard of requirements for forming definitions.

2.2 Step 2: Draw up criteria for definitions

- Draw up a list of criteria for definitions.

2.3 Step 3: Draw up an overview of definitions

- Draw up a list of dimensions and its sources.

2.4 Step 4: Draw up definitions of data quality and evaluate them

- Draw up a section per dimension. In these sections the definitions are copied from the sources.
- Set out the definitions in a table with the following columns:
  - Number: The sequence number refers to a definition.
  - Dimension: For example, the dimension accuracy.
  - Data concept. The data concept associated with the dimension. See section 2.2
  - Definition. The definition of the combination of dimension and data concept.
  - Source: Reference the source as mentioned in Appendix 1.
  - Criteria: Indicates which criteria for definitions are not met.
- The Note section provides explanations and comments on the definitions.

For each dimension, the associated data concept is made explicit in accordance with the Object-Oriented Quality and Risk Management model (Van Nederpelt, 2012). This is relevant because dimensions can be associated with more than one data concept.

For example, a distinction can be made between completeness of data values and completeness of the records. In this case, 'data values' and 'records' are the data concepts. The concept 'data value', for example, also has the dimension 'accuracy'.

The dimension and the data concept must be a logical whole. It must be possible to put the word 'of' between the dimension and the data concept:
<dimension> of **<data concept>**. Example: accuracy of data values

If two concepts are involved, the construction is:
 < dimension> between **<data concept 1>** and **<data concept 2>**.
Example: consistency between data file 1 and data file 2

Generally, the data concept forms part of the definition.

Furthermore:
- Evaluate definitions against the criteria for definitions.
- If definitions meet the requirements, they will be selected as the preferred definition.
- If none of the definitions meet the criteria, a new preferred definition is drawn up using the text of existing definitions or the quality dimension is rejected, because

for example the concept is out of scope. The reason for rejection is mentioned in the Note.

2.5 Step 5: Summarize the definitions

- Collect the preferred definitions.
- Select essential quality dimensions.

2.6 Step 6: Classify the dimensions

- Classify the dimensions by data concept.

3. Results

3.1 Result step 1: Collect sources

All sources are mentions in Appendix 1. In Appendix 2 and 3 the concept system of dimensions of data quality is taken from the report (Black, Van Nederpelt, 2020).

3.2 Result step 2: Draw up criteria for definitions

Criteria for definitions of dimensions of data quality are drawn up in Appendix 4.

3.3 Result step 3: Draw up an overview of definitions

In Appendix 5 a list of dimensions and their sources is drawn up.

3.4 Result step 4: Draw up definitions of data quality and evaluate them

Definitions of dimensions of data quality are collected and evaluated in Appendix 6.

3.5 Result step 5: Summarize the results

The result of the research is a list of the dimensions with their preferred definitions as shown in Table 1.

The most common dimensions (12) are marked blue.

Table 1: 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.
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.

Nr	Dimension	Data Concept	Definition
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.		Data values	The degree to which data values of two sets of attributes ▪ within a record, ▪ within a data file, ▪ between data files, ▪ within a record at different points in time comply with a rule.
21.	Consistency	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.
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.

Nr	Dimension	Data Concept	Definition
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.
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.

Nr	Dimension	Data Concept	Definition
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.
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.

3.6 Result step 6: Classify the dimensions

In Table 2 the dimensions of data quality are classified by data concept. The following data concepts are distinguished:
1. Attribute
2. Data
3. Data file
4. Data value
5. Data value, initial
6. Data pattern
7. Dataset
8. Dataset availability
9. Dataset composition
10. Format
11. Metadata
12. Object (in real-world)
13. Record

Table 2: Definitions of dimensions of data quality classified by data concept

Nr	Data concept	Dimension	Definition
	Attribute		
1.	Attributes	Completeness	The degree to which all required attributes in the dataset are present.
2.	Attributes	Equivalence	The degree to which attributes stored in multiple datasets are conceptually equal.
3.	Attributes	Granularity	The degree to which a single characteristic is subdivided in attributes.
	Data		
4.	Data	Accessibility	The ease with which data can be consulted or retrieved,
5.	Data	Availability	The degree to which data can be consulted or retrieved by data consumers or a process.
6.	Data	Compliance with laws, regulations, or standards	The degree to which data is in accordance with laws, regulations or standards.
7.	Data	Confidentiality	The degree to which disclosure of data should be restricted to authorized data consumers.
8.	Data	Interpretability	The degree to which data are in an appropriate language and units of measure.
9.	Data	Latency	The period of time between the point when the data is created and the point when it is available for use.

Nr	Data concept	Dimension	Definition
10.	Data	Portability	The degree to which data can be installed, replaced or moved from one system to another while preserving the existing quality.
11.	Data	Redundancy	The degree to which logically identical data are stored more than once.
12.	Data	Reputation	The degree to which data are trusted or highly regarded in terms of their source or content.
13.	Data	Traceability	The degree to which data lineage is available.
14.	Data	Obtainability	The degree to which data can be acquired.
15.	Data	Value	The degree to which data provide advantages from their use.
16.	Data	Variety	The degree to which data are available from different data sources.
	Data files		
17.	Data files	Completeness	The degree to which all required data files are present.
18.	Data files	Linkability	The degree to which records of one data file can be correctly coupled with records of another data file.
19.	Data files	Referential integrity	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.
	Data pattern		
20.	Data pattern	Reasonability	The degree to which a data pattern meets expectations.
	Dataset		
21.	Datasets	Access security	The degree to which access to datasets is restricted.
22.	Datasets	Recoverability	The degree to which datasets are preserved in the event of incident.
23.	Dataset	Reproducibility	The degree to which a dataset can be recreated with the same data values.
24.	Datasets	Retention period	The period that datasets are available until they can or must be deleted.
	Dataset availability		

Nr	Data concept	Dimension	Definition
25.	Dataset availability	Punctuality	The degree to which the period between the actual and target point of time of availability of a dataset is appropriate.
26.	Dataset availability	Timeliness	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.
	Dataset composition		
27.	Composition of datasets	Coherence	The degree to which datasets can be combined.
28.	Composition of datasets	Compliance to laws, regulations, or standards	The degree to which the composition of datasets is in accordance with laws, regulations or standards.
29.	Composition of datasets	Naturalness	The degree to which the composition of datasets is aligned with the real-world objects that they represent.
30.	Composition of datasets	Relevance	The degree to which the composition of datasets meets the needs of the data consumer.
	Data value		
31.	Data values	Accuracy	The degree of closeness of data values to real values.
32.	Data values	Comparability of populations	The degree to which data values representing two populations have the same definition and are measured in the same way.
33.	Data values	Comparability over time	The degree to which data values over time have the same definition and are measured in the same way.
34.	Data values	Completeness	The degree to which data values are present.
35.	Data values	Consistency	The degree to which data values of two sets of attributes ▪ within a record, ▪ within a data file, ▪ between data files, ▪ within a record at different points in time comply with a rule.
36.	Data values	Credibility	The degree to which data values are regarded as true and believable by data consumers.

Nr	Data concept	Dimension	Definition
37.	Data values	Currency	The degree to which data values are up to date.
38.	Data values	Integrity	The degree of absence of data value loss or corruption.
39.	Data values	Metadata compliance	The degree to which the data values are in accordance with their definition, format specification and value domain.
40.	Data values	Objectivity	The degree to which the data values are created in an unbiased manner.
41.	Data values	Plausibility	The degree to which data values match knowledge of the real world.
42.	Data values	Precision (1)	The degree of accuracy with which data value are recorded or classified.
43.	Data values	Precision (2)	The degree to which the error in data values spreads around zero (in statistics).
44.	Data values	Validity	The degree to which data values comply with rules.
45.	Data values	Volatility	The degree to which data values change over time.
46.	Data values of an attribute	Completeness	The degree to which all required data values of an attribute are present.
47.	Data values of a set of attributes of a dataset at different points in time (temporal consistency)	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.
48.	Data values of two sets of attributes between datasets (across datasets)	Consistency	The degree to which data values of two sets of attributes between datasets comply with a rule.
49.	Data values of two sets of attributes between records (cross record)	Consistency	The degree to which data values of two sets of attributes between records comply with a rule.
50.	Data values of two sets of attributes within a record (record level)	Consistency	The degree to which data values of two sets of attributes within a record comply with a rule.
	Initial data value		
51.	Initial data value	Reliability	The closeness of the initial data value to the subsequent data value.
	Format		

Nr	Data concept	Dimension	Definition
52.	Format	Ability to represent null values	The degree to which a format allows null values in an attribute.
53.	Format	Appropriateness	The degree to which the format is suitable for use.
54.	Format	Portability	The degree to which a format can be applied in a wide range of situations.
	Metadata		
55.	Metadata	Clarity	The ease by which data consumers can understand the metadata of a dataset.
56.	Metadata	Completeness	The degree to which the metadata are fully described.
	Object		
57.	Objects	Uniqueness	The degree to which objects (of the real world) occur only once as a record in a dataset.
	Record		
58.	Records	Completeness	The degree to which all required records in the data file are present.
59.	Records	Granularity	The degree to which objects are aggregated to records.
60.	Records	Uniqueness	The degree to which records occur only once in a data file.

4.   Conclusions and discussion

The research question was what definitions of dimensions of data quality are preferable and which dimensions are essential. These questions are answered in paragraph 3.1 and 3.2.

Out of 127 definitions from nine authoritative sources, 60 preferred definitions of essential quality dimensions and associated concepts have been drawn up.

Possible follow-up actions are:
a.   Establish a DAMA standard for quality dimensions of data in a separate, more concise report. Or integrate the definitions in DMBOK3.
b.   Enrich each dimension with extra information, e.g., examples, possible quality indicators, possible causes of problems, possible measure to control the quality dimension.
c.   Draw up definitions for quality dimensions for other concepts systems, e.g., databases, statistical, or other concept e.g., data governance, data processing processes, documentation, data sources/suppliers, data presentation.

Acknowledgement

Appendix 1: 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* (2nd 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.

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

McGilvray, D. (2008). *Executing Data Quality Projects: Ten Steps to Quality Data and Trusted Information*. Morgan Kaufman Publishers.

Merriam-Webster. (2020, May 7). Dictionary by Merriam-Webster: America's most-trusted online dictionary. Retrieved from https://www.merriam-webster.com/dictionary

Nederpelt, P.W.M. van (2012). *Object-oriented Quality and Risk Management. A practical method for quality and risk management.* New York/Alphen den Rijn: Lulu/MicroData.

Oxford. (2020, May 7). Oxford learner's dictionaries | Find definitions, translations, and grammar explanations at Oxford learner's dictionaries. Retrieved from https://www.oxfordlearnersdictionaries.com

Regoniel, P. (2015, January 5). Conceptual framework: A step-by-step guide on how to make one. Retrieved from https://simplyeducate.me/2015/01/05/conceptual-framework-guide/

UN/Edifact glossary. (n.d.). Retrieved from https://www.unece.org/trade/untdid/texts/d300_s.htm

Appendix 2: Concept system dimensions of data quality

In this appendix concepts are defined that are relevant in this report. See Table 3. These concepts are taken from the report Concept System of Dimensions of Data Quality (Black, Van Nederpelt, 2020).

A distinction is made between data concepts in de 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 3. This way the coherence between the concepts are made visible.

Table 3: Definitions of concepts

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

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

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

Appendix 3: Diagrams

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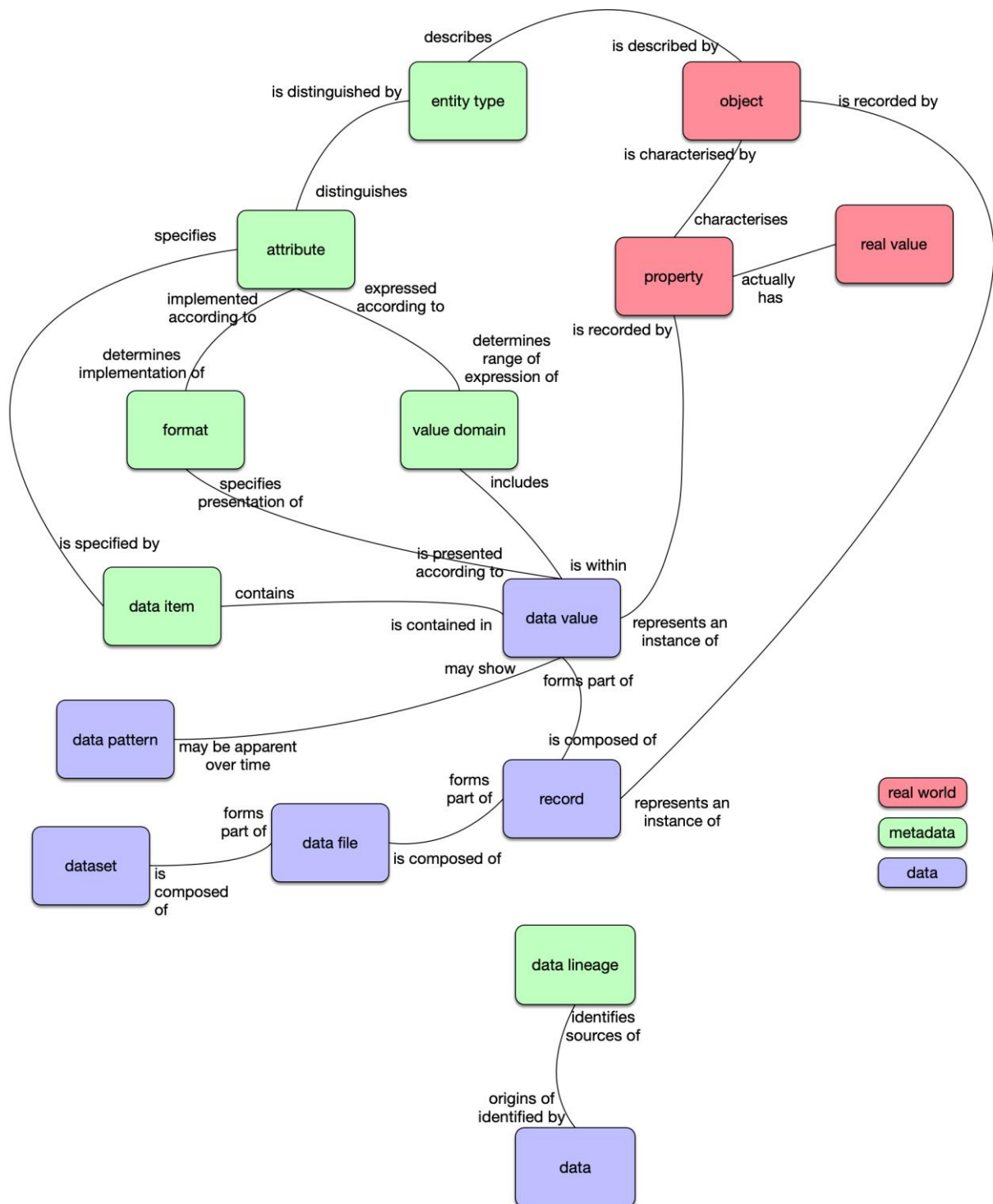
Figure 1: Data concept system DDQ

A set of data concepts in de data world is shown in a data model in Figure 2.

Entity Type

Attribute

Data file

Record

Name	Nationality	Height in cm		
Johnson	GB	180		

Data Value

Data Item
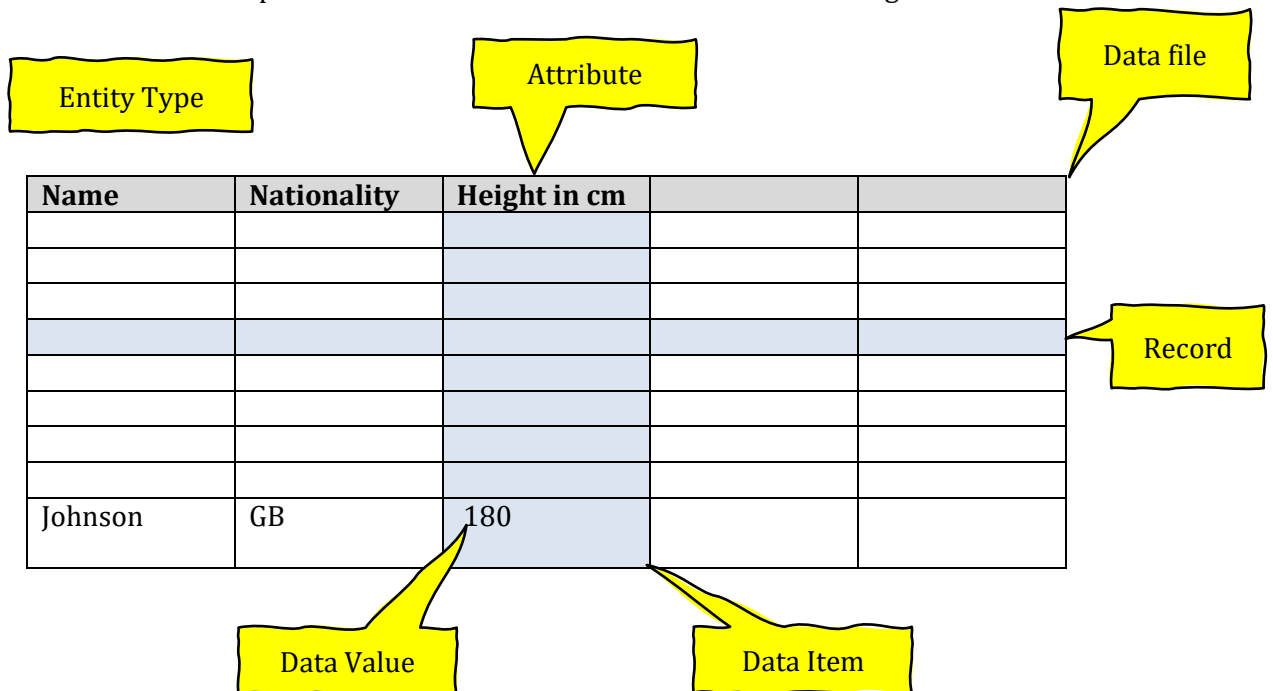
Figure 2: Data concepts in a data model

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

Figure 3: Relationship between data concepts and dimensions

Figure 4 is an artitst impression of the real world and data world.
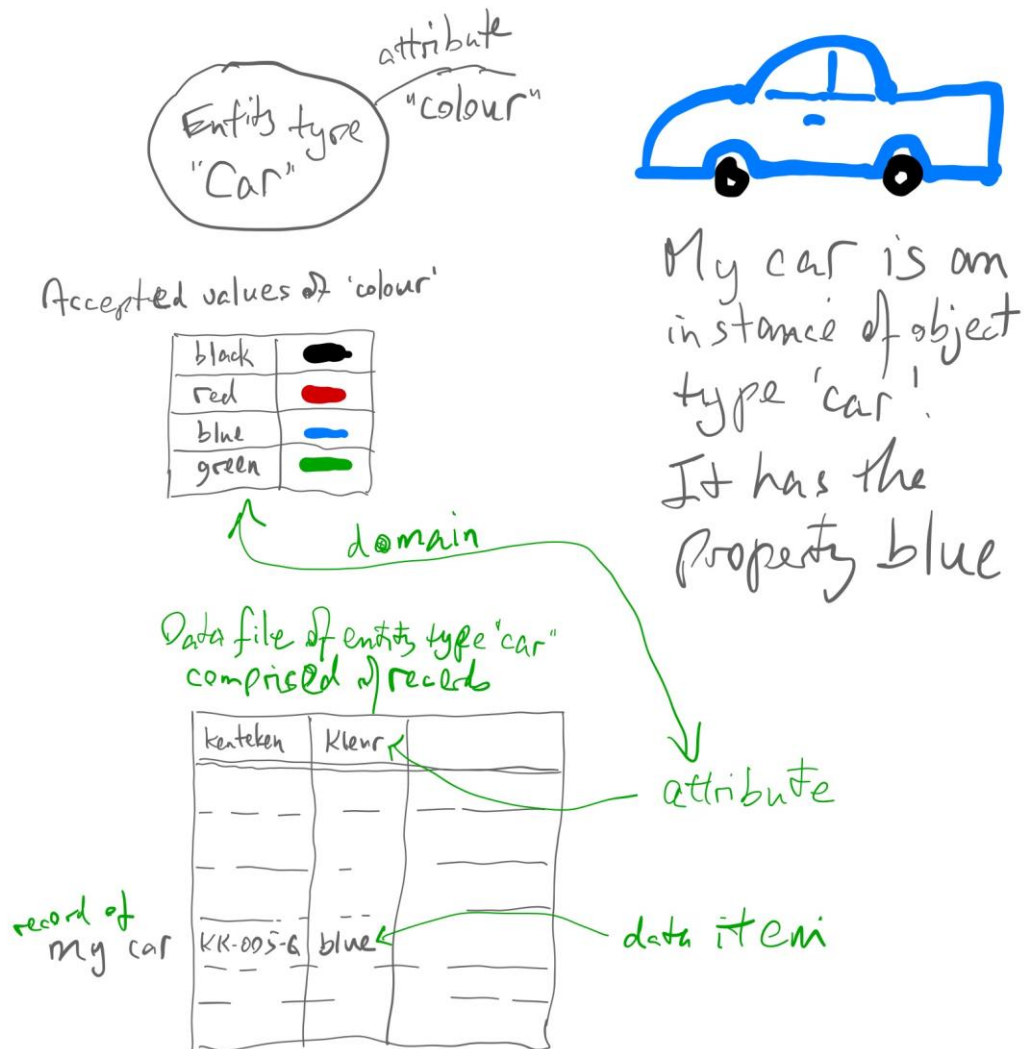
Figure 4: Artists impression of the real world and data world

Appendix 4: Criteria for definitions

The definitions have to comply with the criteria as stated in Table 4. Criteria A until J are laid down in ISO 704 (2009). Criteria K until O are added in order to harmonize all definitions.

Table 4: Criteria for definitions

Code	Section in ISO 704	Criterium
A	6.2	Preferably formulated **'intensional'**, not as an exhaustive list of possibilities.
B	6.2.1	Starting from the **above concept**.
C	6.3.1	The word **'is'** can be put in front of it.
D	6.3.2	Describes the **associations** with other concepts within the same concept system if possible
E	6.3.3	Is as **short** as possible and as complex as necessary.
F	6.3.4	Can replace the term in a sentence (**substitution**).
G	6.4.1	Is not **circular**.
H	6.4.2	Is **not too wide and not too narrow.**
I	6.4.3	Is formulated **positive** (not negative)
J	6.5	There are no additional **comments** in it. They're in a Note.
Code		**Criterium**
K		The definition of a dimension starts with: '**The degree to which'.** Not 'refers to', 'the extent to which', 'measures', etc. It is not formulated as a requirement.
L		If applicable, **concepts** from the real-world, the logical and quality dimension concept system are used in the definition. See Appendix 3.
M		The dimension is **noun** derived from an adverb (accuracy); not the adverb (accurate).
N		Is **readable, unambiguous,** and **written in proper language.**
O		Corresponds with a generally **accepted understanding** of the meaning the quality dimension.

Appendix 5: Dimension of data quality and sources

This Appendix shows a list of dimensions of data quality, their classification assigned by the source and a reference to the source. See Table 5.

Table 5: Dimensions of data quality, classification, and sources

Nr	Dimensions of data quality	Classification	Source
1.	Ease of operation		
2.	Reproducibility		
3.	Granularity		-
4.	Retention period		-
5.	Accessibility		CDDQ 2019
6.	Accuracy		CDDQ 2019
7.	Completeness		CDDQ 2019
8.	Consistency		CDDQ 2019
9.	Currency		CDDQ 2019
10.	Integrity		CDDQ 2019
11.	Lineage		CDDQ 2019
12.	Precision		CDDQ 2019
13.	Representation		CDDQ 2019
14.	Timeliness		CDDQ 2019
15.	Validity		CDDQ 2019
16.	Coverage		Daas 2010
17.	Likability		Daas 2010
18.	Accuracy		DAMA 2017
19.	Completeness		DAMA 2017
20.	Consistency		DAMA 2017
21.	Currency (of data)	Timeliness	DAMA 2017
22.	Integrity or Coherence		DAMA 2017
23.	Latency	Timeliness	DAMA 2017
24.	Reasonability		DAMA 2017
25.	Timeliness		DAMA 2017
26.	Uniqueness		DAMA 2017
27.	Validity		DAMA 2017
28.	Volatility	Timeliness	DAMA 2017
29.	Accuracy	Core dimension	DAMA-UK 2013
30.	Completeness	Core dimension	DAMA-UK 2013
31.	Confidence		DAMA-UK 2013
32.	Consistency	Core dimension	DAMA-UK 2013
33.	Flexibility		DAMA-UK 2013
34.	Timeliness	Core dimension	DAMA-UK 2013
35.	Uniqueness	Core dimension	DAMA-UK 2013
36.	Usability		DAMA-UK 2013

Nr	Dimensions of data quality	Classification	Source
37.	Validity	Core dimension	DAMA-UK 2013
38.	Value		DAMA-UK 2013
39.	Accessibility	Pragmatic	English 1999
40.	Accuracy		English 1999
41.	Accuracy to a surrogate source	Inherent	English 1999
42.	Completeness (of values)	Inherent	English 1999
43.	Concurrency (of redundant or distributed data)	Inherent	English 1999
44.	Contextual clarity	Pragmatic	English 1999
45.	Database integrity		English 1999
46.	Definition conformance (see metadata conformance)	Inherent	English 1999
47.	Entity integrity		English 1999
48.	Equivalence	Inherent	English 1999
49.	Fact completeness	Pragmatic	English 1999
50.	Flexibility		English 1999
51.	Non-duplicates	Inherent	English 1999
52.	Precision	Inherent	English 1999
53.	Stability		English 1999
54.	Timeliness	Pragmatic	English 1999
55.	Usability	Pragmatic	English 1999
56.	Validity	Inherent	English 1999
57.	Accessibility		Eurostat 2015
58.	Accuracy		Eurostat 2015
59.	Clarity		Eurostat 2015
60.	Coherence		Eurostat 2015
61.	Comparability		Eurostat 2015
62.	Confidentiality		Eurostat 2015
63.	Consistency		Eurostat 2015
64.	Punctuality		Eurostat 2015
65.	Relevance		Eurostat 2015
66.	Reliability		Eurostat 2015
67.	Timeliness		Eurostat 2015
68.	Accessibility	Inherent/System dependant	ISO 25012
69.	Accuracy	Inherent	ISO 25012
70.	Availability	System dependant	ISO 25012
71.	Completeness	Inherent	ISO 25012
72.	Compliance	Inherent/System dependant	ISO 25012
73.	Confidentiality	Inherent/System dependant	ISO 25012
74.	Consistency	Inherent	ISO 25012
75.	Credibility	Inherent	ISO 25012
76.	Currentness	Inherent	ISO 25012
77.	Efficiency	Inherent/System dependant	ISO 25012
78.	Portability	System dependant	ISO 25012
79.	Precision	Inherent/System dependant	ISO 25012
80.	Recoverability	System dependant	ISO 25012

Nr	Dimensions of data quality	Classification	Source
81.	Traceability	Inherent/System dependant	ISO 25012
82.	Understandability	Inherent/System dependant	ISO 25012
83.	Ability to represent null values	Representation	Redman 1996
84.	Accordance with format (of the physical instances)	Representation	Redman 1996
85.	Accuracy	Data values	Redman 1996
86.	Appropriateness	Representation	Redman 1996
87.	Clarity	Content	Redman 1996
88.	Completeness	Data values	Redman 1996
89.	Consistency	Data values	Redman 1996
90.	Currency	Data values	Redman 1996
91.	Efficient use (of storage)	Representation	Redman 1996
92.	Flexibility	Reaction to change	Redman 1996
93.	Format flexibility	Representation	Redman 1996
94.	Format precision	Representation	Redman 1996
95.	Granularity (of attributes)	Level of detail	Redman 1996
96.	Homogeneity	Composition	Redman 1996
97.	Identify-ability	Composition	Redman 1996
98.	Interpretability	Representation	Redman 1996
99.	Level of detail	Level of detail	Redman 1996
100.	Naturalness	Composition	Redman 1996
101.	Obtainability	Content	Redman 1996
102.	Portability	Representation	Redman 1996
103.	Precision (of attribute domains)	Level of detail	Redman 1996
104.	Redundancy (minimum necessary)	Composition	Redman 1996
105.	Relevance	Content	Redman 1996
106.	Robustness	Reaction to change	Redman 1996
107.	Scope	Scope	Redman 1996
108.	Semantic consistency (of the components of the model)	View consistency	Redman 1996
109.	Structural consistency (of attributes across entity types)	View consistency	Redman 1996
110.	Access security	Accessibility	Wang 1996
111.	Accessibility	Accessibility	Wang 1996
112.	Accuracy	Intrinsic	Wang 1996
113.	Appropriateness (of amount of data)	Contextual	Wang 1996
114.	Believability	Intrinsic	Wang 1996
115.	Completeness	Contextual	Wang 1996
116.	Conciseness (of representation)	Representational	Wang 1996
117.	Cost-effectiveness		Wang 1996
118.	Ease of understanding	Representational	Wang 1996
119.	Interpretability	Representational	Wang 1996
120.	Objectivity	Intrinsic	Wang 1996
121.	Relevancy	Contextual	Wang 1996
122.	Representational consistency	Representational	Wang 1996
123.	Reputation	Intrinsic	Wang 1996
124.	Timeliness	Contextual	Wang 1996
125.	Traceability		Wang 1996
126.	Value-added	Contextual	Wang 1996
127.	Variety		Wang 1996

Appendix 6: Definitions of dimensions of data quality

This appendix defines the quality dimensions associated with various data concepts.

Caption

* = Derived from source.

The column 'Criteria' indicates with a letter code from Table 4 which criteria for definitions are not met.

Preferred definitions are marked blue

1. Ability

1.1 Definitions

Nr	Dimension	Data concept	Definition	Source	Criteria not met
1	Ability to represent null values	Formats	Not available	Redman 1996	C, F, K, L
2	Ability to represent null values	Format	The degree to which a format allows null values in an attribute.	-	None

1.2 Notes

-

2. Access security

2.1 Definitions

Nr	Dimension	Data concept	Definition	Source	Criteria not met
1	Access security	Data	The extent to which access to data can be restricted and hence kept secure.	Wang 1996	E, G, K
2	Access security	Datasets	The degree to which access to datasets is restricted.	Wang 1996*	

2.2 Notes

-

3.   Accessibility

3.1   Definitions

Nr	Dimension	Data concept	Definition	Source	Criteria not met
1	Accessibility	Data	Measures how easy it is to acquire data when needed, how long it retained, and how access is controlled.	CDDQ 2019	H, K, L
2	Accessibility and clarity	Statistic	Refer to the simplicity and ease, the conditions and modalities by which users can access, use and interpret statistics, with the appropriate supporting information and assistance.	Eurostat 2015	H, K, L
3	Accessibility	Data	The extent to which data are available or easily and quickly retrievable.	Wang 1996	E, H
4	Accessibility	Data	The characteristic of being able to access data when it is required.	English 1999	K, O
5	Accessibility	Data	The degree to which data can be accessed in a specific context of use, particularly by people who need supporting technology of special configuration because of some disability.	ISO 25012	E, H, J, O
6	Accessibility	Data	The ease with which data can be consulted or retrieved.		

3.2   Notes
- A prerequisite for accessibility is availability.
- Definition 4: See also availability.

4. Accuracy

4.1 Definitions

Nr	Dimension	Data concept	Definition	Source	Criteria not met
1	Accuracy	Data	The degree to which data factually represent its associated real-world object, event.	CDDQ 2019	L
2	Accuracy	Data	The degree to which data correctly describes the "real world" object or event being described	DAMA-UK 2013	L
3	Accuracy	Data value	The degree to which data values correctly represents real-world entities.	DAMA 2017	L
4	Accuracy	Statistical output	The degree of closeness of computations or estimates to the exact or true values that the statistics were intended to measure.	Eurostat 2015	L
5	Accuracy	Data	The extent to which data are correct, reliable, and certified free of error.	Wang 1996	E, K, L
6	Accuracy to reality	Data value	A characteristic of information quality measuring the degree to which a data value (or set of data values) correctly represents the attributes of the real-world object or event.	English 1999	E, K
7	Accuracy to surrogate source	Data	A measure of the degree to which data agrees with an original, acknowledged authorative source of data about a real world object or event, such as a form, document, or unaltered electronic data received from outside the organization.	English 1999	E, J, K, L
8	Accuracy	Datum (e, a, v)	Refers to the nearness of the value v to some value v' in the attribute domain, which is considered as the correct on for the entity e and the attribute A.	Redman 1996	C, E, H, K, L.

Nr	Dimension	Data concept	Definition	Source	Criteria not met
9	Accuracy	Data	The degree to which data has attributes that correctly represent the true value of the intended attribute of a concept or event in a specific context of use.	ISO 25012	E, K, O
10	Syntactic accuracy	Data values	The closeness of the data values to a set of values defined in a domain considered *syntactically* correct.	ISO 25012	K, O
11	Semantic accuracy	Data values	The closeness of the data values to a set of values defined in a domain considered *semantically* correct.	ISO 25012	K, O
12	Accuracy	Data item and entity	Agreement between a data item and the entity that it represent. For reference, accuracy should be checked to assure that: each data item links to a specific entity; each entity has a data entry related to it.	King 2020	E, J, N
13	Accuracy	Data values	The degree of correspondence between data values to real values.		None

4.2 Notes

- Definition 9, 10 and 11: See validity and metadata conformity.

5. Appropriateness

5.1 Definitions

Nr	Dimension	Data concept	Definition	Source	Criteria not met
1	Appropriate	Amount of data	The extent to which the quantity or volume of available data is appropriate	Wang 1996	E, G, K, M
2	Appropriateness	Format	Depends on user and medium.	Redman 1996	C, F, K, L, N
3	Appropriateness	Format	The degree to which the format is suitable for use.	-	None

5.2 Notes

- Definition 1: See relevance. It depends on the composition of a dataset, what the amount of data is.
- Appropriateness can be associated with a lot of concepts: data, metadata, attributes, records, datasets, documentation, etc. The added value of this quality dimension is small.

6. Availability

6.1 Definitions

Nr	Dimension	Data concept	Definition	Source	Criteria not met
1	Availability	System or data	A percentage measure of the reliability of a system indicating the percentage of time the system or data is accessible or usable, compared to the amount of time the system or data should be accessible or usable.	English 1999	K, L
2	Availability	Data	The degree to which data has attributes that enable it to be retrieved by authorized users and/or applications in a specific context of use.	ISO 25012	H,
3	Availability	Data	The degree to which data can be consulted or retrieved by data consumers or processes.		

6.2 Notes

- Accessibility is about the ease with which data can be consulted or retrieved. Availability says if it can be retrieved at all.
- Availability can be specific for a group of people or processes, e.g., a dataset is not available to the public but only for some authorized people.
- There is a relationship with access security.

7. Believability

7.1 Definitions

Nr	Dimension	Data concept	Definition	Source	Criteria not met
1	Believability		Not available.	Wang 1996	

7.2 Notes

- See credibility.

8. Clarity

8.1 Definitions

Nr	Dimension	Data concept	Definition	Source	Criteria not met
1	Accessibility and clarity	Statistics	Refer to the simplicity and ease, the conditions and modalities by which users can access, use and interpret statistics, with the appropriate supporting information and assistance.	Eurostat, 2014	H, K, L
2	Clarity	Context	The relative degree to which data presentation enables the knowledge worker to understand the meaning of the data and avoid interpretation.	English 1999	E, K, L
3	Clarity	Definitions	Not available	Redman 1996	C, F, K, L
4	Clarity	Metadata	The ease with which data consumers can understand the metadata.		

8.2 Notes

- Data can only be interpreted by reference to the associated metadata.
- Clarity has two subdimensions: readability and unambiguity.

9. Coherence

9.1 Definitions

Nr	Dimension	Data concept	Definition	Source	Criteria not met
1	Coherence	Statistics	Measures the adequacy of the statistics to be combined in different ways and for various uses.	Eurostat 2015	K, L
2	Coherence	Composition of datasets	The degree to which datasets can be combined.		

9.2 Notes

- Datasets can be combined if there are enough similarities in the composition of a dataset, such as same object type, same reference time, same classification system in case of aggregated data.

10. Comparability

10.1 Definitions

Nr	Dimension	Data concept	Definition	Source	Criteria not met
1	Comparability	Data	Measurement of the impact of differences in applied statistical concepts, measurement tools and procedures where statistics are compared between geographical areas or over time.	Eurostat, 2014	H, K, L
2	Comparability over time	Data values	The degree to which data values over time have the same definition and are measured in the same way.		
3	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.		

10.2 Note

- Examples of comparability over time: employment rate in 2020 can be compared with employment rate in 2019.
- Example of comparability between data values of two datasets: employment rate in Europe can be compared with employment rate in the USA.

11. Completeness

11.1 Definitions

Nr	Dimension	Data concept	Definition	Source	Criteria not met
1	Completeness	Data	Refers to whether all required data is present.	DAMA 2017	H, K, L
2	Completeness	Data values	Measures the degree of population of data values in a dataset.	CDDQ 2019	K, L
3	Completeness	Data	The proportion of stored data against the potential of "100% complete"	DAMA-UK 2013	H, K, L
4	Completeness	Data	The extent to which data are of sufficient breadth, depth, and scope for the task at hand.	Wang 1996	K, N
5	Completeness	Data	The degree to which all required data is known.	English 1999	
6	Fact completeness	Attributes	Percentage of the attributes about an entity type that need to be known to assure that they are defined to assure that they are defined in the model and implemented in a data base.	English 1999	K, L
7	Value completeness	Data content	Percentage of the columns or fields of a table or file that should have values in them, in fact do so. Also referred to as coverage.	English 1999	K, L
8	Completeness	Values	The degree to which values are present in a data collection.	Redman 1996	K, L
9	Completeness	Entities		Redman 1996	C, E, F, K, L
10	Completeness	Attributes		Redman 1996	C, E, F, K, L
11	Completeness	Subject data	The degree to which subject data associated with an entity has values for all expected attributes and related entity instances in a specific context of use.	ISO 25012	E, L, N
12	Completeness	Data records	The quality of having data records stored for all entities and that all attributes for an entity are populated.	King 2020	K
13	Completeness	Data values	The degree to which all required data values are present.		
14	Completeness	Records	The degree to which all required records in the dataset are present.		
15	Completeness	Data values of an attribute	The degree to which all required data values of an attribute are present.		

Nr	Dimension	Data concept	Definition	Source	Criteria not met
16	Completeness	Attributes	The degree to which all required attribute in the dataset are present.		
17	Completeness	Data files	The degree to which all required data files are present.		
18	Completeness	Metadata	The degree to which the metadata are fully described.		

11.2 Notes

- Completeness rules can be assigned with varying degrees of constraints:
    - mandatory attributes require a value,
    - conditional and optional attribute are filled according to a rule or dependency.

12. Compliance

12.1 Definitions

Nr	Dimension	Data concept	Definition	Source	Criteria not met
1	Compliance	Data	The degree to which data has attributes that adhere to standards, conventions or regulations in force and similar rules relating to data quality in a specific context of use.	ISO 25012	E
2	Compliance with laws, regulations, or standards	Data	The degree to which data is in accordance with laws, regulations or standards.	-	None
3	Compliance with laws, regulations, or standards	Composition of datasets	The degree to which the composition of datasets is in accordance with laws, regulations or standards.	-	None

12.2 Relationships with other dimensions

- Compliance with laws, regulations or standards is like validity of data values, which is the degree to which data values comply with rules.,

12.3 Notes

-

13. Conciseness

13.1 Definitions

Nr	Dimension	Data concept	Definition	Source	Criteria not met
1	Concise	Data	The extent to which data are compactly represented without being overwhelming (i.e., brief in presentation, yet complete and to the point).	Wang 1996	E, I, J, K, M
2	Conciseness	Data presentation	The degree to which data is presented in a brief but comprehensive way.	-	None

13.2 Notes

- Concise means brief but comprehensive. Comprehensive means including or dealing with all or nearly all elements or aspects of something.
- Other quality dimensions of presentation of data can be distinguished, such as completeness, format, clarity, lay-out, appearance, adequacy.
- Data presentation may be added as a data concept at a later stage.

14. Concurrency

14.1 Definitions

Nr	Dimension	Data concept	Definition	Source	Criteria not met
1	Concurrency	Data	The degree to which the timing of equivalence of data is stored in redundant or distributed database files.	English 1999	L, N
2	Concurrency	Redundant or distributed data.	-	DMBOK2	C, E, F
3	Concurrency	-	The simultaneous execution of processes against the database.	Early 2011	C, F

14.2 Notes

- What is meant by timing of equivalence of data?

15. Confidence

15.1 Definitions

Nr	Dimension	Data concept	Definition	Source	Criteria not met
1	Confidence	Data	Are Data Governance, Data Protection and Data Security in place? What is the reputation of the data, and is it verified or verifiable?	DAMA UK 2013	C, K, L

15.2 Notes

- Can confidence be regarded as a quality dimension? One can have confidence in something (but not of something).
- See quality dimension reputation.

16. Confidentiality

16.1 Definitions

Nr	Dimension	Data concept	Definition	Source	Criteria not met
1	Confidence	Data	Are Data Governance, Data Protection and Data Security in place? What is the reputation of the data, and is it verified or verifiable?	DAMA UK 2013	C, K, L
2	Confidentiality	Data	The extent to which their unauthorized disclosure could be prejudicial or harmful to the interest of the source or other relevant parties.	Eurostat 2014	I, K
3	Confidentiality	Data	The degree to which data has attributes that ensure that it is only accessible and interpretable by authorized users in a specific context of use.	ISO 25012	E
4	Confidentiality	Data	The degree to which disclosure of data should be restricted to authorized data consumers.	-	None

16.2 Notes

-

17. Cost effectiveness

17.1 Definitions

Nr	Dimension	Data concept	Definition	Source	Criteria not met
1	Cost effectiveness	Data accuracy, cost of data collection	The extent to which the cost of collecting appropriate data is reasonable.	Wang 1996	C, F, K, L
2	Cost effectiveness	Data processing process	The degree to which data collecting and processing is economical in terms of tangible benefits produced by money spent.		

17.2 Notes

- Cost-effectiveness means that there is a positive relationship between costs and benefits or efficiency and effectiveness.
- Mariam-Webster defines cost-effective as: economical in terms of tangible benefits produced by money spent.
- Definition 1 and 2: Concept is out of scope of this report.

18. Consistency

18.1 Definitions

Nr	Dimension	Data concept	Definition	Source	Criteria not met
1	Consistency	Data values	Can refer to ensuring that data values are consistently represented within a dataset and between datasets, and consistently associated across datasets.	DAMA 2017	E, C, G, K
2	Consistency	Representation of things	The absence of difference, when comparing two or more representations of a thing against a definition.	DAMA UK 2013	K, L
3	Consistency	Data	Measures whether or not data is equivalent across systems or location of storage.	CDDQ 2019	K, L
4	Consistency	Representations of a thing	The absence of difference, when comparing two or more representations of a thing against a definition.	DAMA-UK 2013	K, L
5	Semantic consistency	Components of the view	Refers to the consistency of the various components of the view.	Redman 1996	C, G, K, L, O
6	Structural consistency	Attributes	Refers to the consistency of attributes across entity types.	Redman 1996	C, G, K, L, O
7	Consistency	Values	Not available	Redman 1996	C, E, K, L, N
8	Consistency	Physical instances of data	Refers to whether physical instances of data are in accord with their format.	Redman 1996	C, F, K, L

Nr	Dimension	Data concept	Definition	Source	Criteria not met
9	Consistency	Data	The degree to which data has attributes that are free from contradiction and are coherent with other data in a specific context of use. It can be either or both among data regarding one entity and across similar data for comparable entities.	ISO 25012	E, H, J, N
10	Consistency	Data	The ability to correctly link data relating to the same entity across different datasets.	King 2020	H, K
11	Consistency	Data values	The degree to which data values of two sets of attributes - within a record, - within a data file, - between data files, - within a record at different points in time comply with a rule.		
12	Consistency	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.		
13	Consistency	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.		
14	Consistency	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.		
15	Consistency	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.		

18.2 Notes

- In consistency two or more data values are compared.
- There are four types of consistency:
    - consistency within a record,
    - consistency between records within a dataset,
    - consistency across datasets and
    - temporal consistency.
- Plausibility of data values can be seen of a type of consistency.
- A rule can be an equation with a =, > or < symbol. For example, profit = return – costs.
- A rule can also be business rule. For example, one might expect that the number of transactions each day does not exceed 105% of the running average number of transactions for the previous 30 days.
- Do not confuse consistency with accuracy or correctness.

19. Coverage

19.1 Definitions

Nr	Dimension	Data concept	Definition	Source	Criteria not met
1	Coverage	Population	Percentage of units not belonging to population or missing from the target population.	Daas 2010	C, F, K, L

19.2 Notes

- Definition 1: See completeness of records.

20. Credibility

20.1 Definitions

Nr	Dimension	Data concept	Definition	Source	Criteria not met
1	Credibility	Data	The degree to which data has attributes that are regarded as true and believable by users in a specific context of use. Credibility includes the concept of authenticity (the truthfulness of origins, attributions, commitments).	ISO 25012	E, J
2	Credibility	Data values	The degree to which data values are regarded as true and believable by data consumers.		

20.2 Notes

- Objectivity can be regarded as a subdimension of credibility; it contributes to credibility.

21. Currency

21.1 Definitions

Nr	Dimension	Data concept	Definition	Source	Criteria not met
1	Currency	Data values	The measure of whether data values are the most up-to-date version of the information.	DAMA 2017	K
2	Currency	Data	Measures how quickly data reflects the real world concept that is represent.	CDDQ 2019	K
3	Currency	Datum	The degree to which a datum in question is up-to-data.	Redman 1996	E, L
4	Currentness	Data	The degree to which data has attributes that are of the right age in a specific context of use.	ISO 25012	E, O
4	Currency	Data values	The degree to which data values are up-to-date	Redman 1996*	

21.2 Notes

- According to DAMA (2017), currency is a subdimension of timeliness.
- Timelines and currency are dependent on each other. The better the timeliness of data values the better currency of data values.
- Definition 2 and 4: See timeliness.

22. Data lineage

22.1 Definitions

Nr	Dimension	Data concept	Definition	Source	Criteria not met
1	Lineage	Data	Measures whether factual documentation exists about where data came from, how it was transformed, where it went and end-to-end graphical illustration.	CDDQ 2019	E, K, L

22.2 Note

- Data lineage is metadata that identifies the sources of data and the transformations through which it has passed up to the point of consumption. Data lineage is therefore not a quality dimension but a data concept.
- Data lineage and traceability of data are related. Data lineage makes traceability of data possible.

23. Ease of understanding

23.1 Definitions

Nr	Dimension	Data concept	Definition	Source	Criteria not met
1	Ease of understanding	Data	The extent to which data are clear without ambiguity and easily comprehended.	Wang 1996	E, K

23.2 Notes

- Definition 1: See clarity.
- Ease of understanding depends on the availability of unambiguous metadata.

24. Ease of operation

24.1 Definitions

Nr	Dimension	Data concept	Definition	Source	Criteria not met
1	Ease of operation	Data	The extent to which data are easily managed and manipulated (i.e., updated, moved, aggregated, reproduced, customized).	Wang 1996	H, K, L

24.2 Notes

- Ease of operation is not dependant of the data itself but on quality of the information system which includes the data, and the format of the dataset.

25. Efficiency

25.1 Definitions

Nr	Dimension	Data concept	Definition	Source	Criteria not met
1	Efficiency	Usage of Recording Media	Not available	Redman 1996	C, F, K, L
2	Efficiency	Data	The degree to which data has attributes that can be processed and provide the expected levels of performance by using the appropriate amounts and types of resources in a specific context of use.	ISO 25012	E, N

25.2 Notes

- Definition 1: Usage of recording media is out of scope of this report.
- Efficiency is generally associated to processes, not to data.
- The volume of data in storage or in communication can vary depending on the methodology or syntax used. E.g. the same data in an XML file can be more voluminous than in a CSV or JSON file. But this is a technical implementation issue rather than a data quality issue.

26. Equivalence

26.1 Definitions

Nr	Dimension	Data concept	Definition	Source	Criteria not met
1	Equivalence	Data	The degree to which data stored in multiple places is conceptually equal.	English 1999 (1)	L
2	Equivalence	Data/Redundant or distributed databases	The degree to which data in one data collection or database is semantically equivalent to data about the same object or event in another data collection or database.	English 1999 (2)	G
3	Equivalence	Attributes	The degree to which attributes stored in multiple datasets are conceptually equal.	English 1999*	

26.2 Notes

- Equivalence indicates that data attributes have equal values or, in essence, is the same. For example, two datasets contain the data attribute 'gender', but the sets use different gender codes. (English, 1999)
- To be fully equivalent, the classification system (e.g., of gender) should be equal. The codes can be different but not the meaning of the codes.
- In the table below, gender code A and gender code B are not fully equivalent.

Gender code A	Gender code B
M	M
F	V
	X

27. Flexibility

27.1 Definitions

Nr	Dimension	Data concept	Definition	Source	Criteria not met
1	Flexibility	Data	Is it comparable and compatible with other data, does it have useful groupings and classifications? Can it be repurposed, and is it easy to manipulate?	DAMA-UK 2013	C, E, F, G, K
2	Flexibility	Data	The extent to which data are expandable, adaptable, and easily applied to other needs.	Wang 1996	K
3	Flexibility	Information architecture or database	The degree to which the information architecture or database is able to support organizational or process reengineering changes with minimal modification of the exiting objects and relationships, only adding new objects and relationships.	English 1999	E, L
4	Flexibility	Capacity to change	Refers to the capacity to change a view to accommodate new demands.	Redman 1996	C, F, K, L, N
5	Flexibility	Format	Changes in user's needs can be accommodated.	Redman 1996	C, F, K, L

27.2 Notes

- Definition 1: includes other quality dimension such as relevancy and comparability.
- Definition 2: It is unclear what is meant. Data can't be flexible. Flexibility seems a quality dimension of other concepts such as a system, a database, storage, processing capacity or the software.
- Definition 4: It is unclear which data concept should be flexible.
- Definition 5: Format includes value domain, datatype, unit of measure and character set. It depends on the flexibility of the database if a format can be accommodated. The database is not part of the concept system.

28. Granularity

28.1 Definitions

Nr	Dimension	Data concept	Definition	Source	Criteria not met
1	Granularity	Data fields	Refers to the size in which data fields are sub-divided	Wikipedia	K, L
2	Granularity	Entities	The extent to which groups of smaller indistinguishable entities have joined together to become larger distinguishable entities.	Wikipedia	K, L
3	Granularity	Attributes	Refers to the number and coverage of attributes that are used to represent a single concept.	Redman 1996	H, K, L
4	Granularity	Attributes	The degree to which a single characteristic is subdivided in attributes.		
5	Granularity	Records	The degree to which objects are aggregated to records.		

28.2 Notes

- For example, characteristic postal address can be recorded, with *coarse granularity*, as a single attribute:
  1. address = 200 2nd Ave. South #358, St. Petersburg, FL 33701-4313 USA
  or with *fine granularity*, as multiple attributes:
  1. street address = 200 2nd Ave. South #358
  2. city = St. Petersburg
  3. state = FL
  4. postal code = 33701-4313
  5. country = USA
- Granularity of records implies that aggregation has taken place, e.g., an aggregation of income per gender.
- Synonym of granularity is level of detail.

29. Homogeneity

29.1 Definitions

Nr	Dimension	Data concept	Definition	Source	Criteria not met
1	Homogeneity		Attributes are generally applicable across as many entities of a given type as possible.	Redman 1996	C, F, K, L, N

29.2 Notes

- Definition1: Is not truly clear.

30. Identifiability

30.1 Definitions

Nr	Dimension	Data concept	Definition	Source	Criteria not met
1	Identifiability		Each entity is easily and uniquely identifiable.	Redman 1996	

30.2 Notes

- Definition 1: See uniqueness of objects.

31. Integrity

31.1 Definitions

Nr	Dimension	Data concept	Definition	Source	Criteria not met
1	Referential integrity	Two datasets	Consistency between data objects via a reference key contained in both objects.	DAMA 2017	K, L
2	Integrity or corruption	Data values	Internal consistency within a dataset such that there are no holes or missing parts.	DAMA 2017	K, L
3	Integrity	Dataset	Measures the structural or relational quality of datasets.	CDDQ 2019	K, L
4	Integrity	Database	The characteristic of data in a database in which the data conform to the physical integrity constraints, such as referential integrity and primary key uniqueness, and is able to be secured and recovered in the event of an application, software, or hardware failure.	English 1999	E, H, K, L,
5	Integrity	Entity	The assurance that a primary key value will identify no more than one occurrence of an entity type, and that no attributes of the primary key may contain a null value.	English 1999	K, L, O
6	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.		
7	Integrity	Data values	The degree of absence of data value loss or corruption.		

31.2 Notes

- Referential integrity can be seen as a specific type of consistency, namely consistency across datasets. In case of referential integrity keys are involved.
- Datasets with referential integrity problems have 'orphans' in one or both datasets. This means than records in one dataset are not related to records the other dataset.
- Problems with referential integrity can occur in the dataset with the primary key as well in the dataset with the foreign key.
- Problems with reference integrity can be absence of records, invalid keys, or missing keys.
- Two definitions of integrity are possible because the concept is different.

32. Interpretability

32.1 Definitions

Nr	Dimension	Data concept	Definition	Source	Criteria not met
1	Interpretability	Format	Not available	Redman 1996	C, E, F, K, L
2	Interpretability	Data	The extent to which data are in an appropriate language and units and the data definitions are clear.	Wang 1996	K, N
3	Interpretability	Data	The degree to which data are in an appropriate language and units of measure.	-	None

32.2 Relationships with other dimensions
- The interpretability of data is dependent on the appropriateness of the format (character set, unit of measurement).

32.3 Notes
- Example: It is appropriate to record addresses in Thailand in Thai characters. It is appropriate to record length in GB in inches and not in centimetres. Date formats should be clear too: DD MM YY or MM DD YY.

33. Latency

33.1 Definitions

Nr	Dimension	Data concept	Definition	Source	Criteria not met
1	Latency	Data availability	The time between when the data was created and when it was made available for use.	DAMA 2017	K
2	Latency	Data availability	The period of time between the point when the data is created and the point when it is available for use.	-	K

33.2 Relationship with other dimensions

- Latency is related to timeliness. However, timeliness (Eurostat, 2014) starts at an earlier point in time, i.e., the time that the event or phenomenon took place:

Time that the event or phenomenon takes place	Time of data creation	Time of availability of data for data consumers
Timeliness starts here	Latency starts here	

- Latency is in general a short period of time, e.g. a few milliseconds. Timeliness can take even a few years. Latency has to do with processing delay in the system.

33.3 Note

-

34. Level of detail

34.1 Definitions

Nr	Dimension	Data concept	Definition	Source	Criteria not met
1	Level of detail	Data	Refers to the quantity of data to be included and how precise those data must be.	Redman 1996	C, K

34.2 Relationships with other dimensions

- Level of detail is subdivided to
  1) granularity of attributes and
  2) precision of domains (Redman 1996) e.g. dimensions in cm or μm.

35. Linkability

35.1 Definitions

Nr	Dimension	Data concept	Definition	Source	Criteria not met
1	Linkability	Units	Percentage of units linked unambiguously. Percentage of units incorrectly linked.	Daas 2010	C, E, F, I, K, L
2	Linkability	Data files	The degree to which records of one data file can be correctly coupled with records of another data file.		

35.2 Notes

- Coherence is a prerequisite for linkability.
- Linking is generally used to enrich a dataset with data values of another dataset.

36. Metadata compliance

36.1 Definitions

Nr	Dimension	Data concept	Definition	Source	Criteria not met
1	Metadata compliance	Data	Does the data values comply to the metadata definitions?	Daas 2010	C, F, K, L
2	Definition conformance	Data	The characteristic of data, such that the data values represent a fact consistent with the agreed-upon definition of the attribute.	English 1999	E, K
3	Definition conformance	Actual data values	The consistency of the meaning of the actual data values with its data definition.	English 1999 p143	E, K
3	Metadata compliance	Data values	The degree to which the data values are in accordance with their definition, format specification and value domain.		

36.2 Notes

-

37. Naturalness

37.1 Definitions

Nr	Dimension	Data concept	Definition	Source	Criteria not met
1	Naturalness	Entity types, attributes, domains	Encompass a number of desirable characteristics, such as the use of single-fact data.	Redman 1996	C, F, K, L
2	Naturalness	Composition of datasets	The degree to which the composition of datasets is aligned with the real-world objects that they represent.		

37.2 Notes

-

38. Objectivity

38.1 Definitions

Nr	Dimension	Data concept	Definition	Source	Criteria not met
1	Objectivity	Data	The extent to which data are unbiased (unprejudiced) and impartial.	Wang 1996	K, N
2	Objectivity	Data values	The degree to which the data values are created in an unbiased manner.		

38.2 Note

- Bias is a subdimension of accuracy.
- Biased data values are not accurate but not all inaccurate data values are biased.

39. Obtainability

39.1 Definitions

Nr	Dimension	Data concept	Definition	Source	Criteria not met
1	Obtainability	Values	Not available.	Redman 1996	All
2	Obtainability	Data	The degree to which data can be acquired.		

39.2 Notes

-

40. Plausibility

40.1 Definitions

Nr	Dimension	Data concept	Definition	Source	Criteria not met
1	Plausibility	Data values	The degree to which data values match knowledge of the real world.		

40.2 Notes

- Knowledge of the world can be, e.g., information in a newspaper, data from other sources.
- Plausibility can be seen as a category of consistency, i.e., consistency between the data values and knowledge of the world.

41. Portability

41.1 Definitions

Nr	Dimension	Data concept	Definition	Source	Criteria not met
1	Portability	Formats	Means that they can be applied to as wide a range of situations as possible.	Redman 1996	C, F, K, L
2	Portability	Data	The degree to which data has attributes that enable it to be installed, replaced or moved from one system to another (while) preserving the existing quality in a specific context of use.	ISO 25012	E
3	Portability	Format	The degree to which a format can be applied in a wide range of situations.	Redman 1996*	None
3	Portability	Data	The degree to which data can be installed, replaced or moved from one system to another while preserving the existing quality.	ISO 25012*	None

41.2 Notes

-

42. Precision

42.1 Definitions

Nr	Dimension	Data concept	Definition	Source	Criteria not met
1	Precision	Classification detail	Measurement or classification detail used in specifying an attribute's domain.	CDDQ 2019	N, L
2	Precision	Data	The degree to which data is known to the right level of granularity.	English 1999	L
3	Precision	Estimate	The degree to which the error in the estimate spreads around zero.	Van Nederpelt (2009)	L
4	Precision	Domain	Refers to the level of detail in the measurement or classification scheme that defines the domain.	Redman 1996	L
5	Precision	Classification	Makes clear why icons and colors are of limited use when domains are large.	Redman 1996	L
6	Precision	Data	The degree to which data has attributes that are exact or that provide discrimination in a specific context of use.	ISO 25012	N
7	Precision	Data entry	Degree of specificity for a data entry (ISO 11179)	In: King 2020	None
8	Precision (1)	Data value	The degree of accuracy with which the data value is recorded or classified.		
9	Precision (2)	Data value	The degree to which the error in the data value spreads around zero (in statistics).		

42.2 Notes

- Definition 1: Looks equivalent to granularity.
- Definition 2: For example, a percentage value with two decimal points (0.00%) discriminates to the closest $1/100^{th}$ of a percent (English 1999).
- Definition 3: Synonyms are variance or random error.

88

- Definition 4: Redman (1996) explains precision of domains with the following example. Height of a person, as measured in inches, provides for greater precision than height measured in feet. Similarly, a classification scheme involving 20 categories is more precise than one involving seven categories.
- Definition 5: Format is out of scope of this report.

43. Punctuality

43.1 Definitions

Nr	Dimension	Data concept	Definition	Source	Criteria not met
1	Punctuality	Release time of data	The time lag between the actual delivery of data and the target date on which they were scheduled for release as announced in an official release calendar, laid down by Regulations or previously agreed among partners.	Eurostat, 2014	E
2	Punctuality	Dataset availability	The degree to which the period between the actual and target point in time of availability of a dataset is appropriate.		

43.2 Notes

- A dataset can be made available in a passive way (pull) or active way (push) to data consumers.
- Target time can be planned in advance or agreed.
- A dataset can be made available too early or too late.

44. Reasonability

44.1 Definitions

Nr	Dimension	Data concept	Definition	Source	Criteria not met
1	Reasonability	Data pattern	Asks whether a data pattern meets expectations.	DAMA 2017	K
2	Reasonability	Data pattern	The degree to which a data pattern meets expectations.	DAMA 2017*	None

44.2 Notes

- Examples of data patterns and expectations

Data pattern	Expectation
Number of customers distributed over regions	Sales = number of customers x A
Any data	Benchmark
Sales 2020	Is higher than sales 2019
Time series	Follows trend

- Four types of trends can be distinguished in time series:
  - Trends with following subtypes:
    - Linear Trend
    - Exponential Trend
    - Damped Trend
  - Seasonal
  - Cyclical
  - Random or Irregular or Error

  (What Are Data Trends and Patterns, 2018)
- Expectations comes from knowledge of the real world, goals, other sources or are derived from the data values itself.

45. Recoverability

45.1 Definitions

Nr	Dimension	Data concept	Definition	Source	Criteria not met
1	Recoverability	Data	The degree to which data has attributes that enable it to maintain and preserve a specified level of operations and quality, even in the event of failure, in a specific context of use.	ISO 25012	E
2	Recoverability	Datasets	The degree to which datasets are preserved in the event of incident.	ISO 25012*	

45.2 Notes

-

46. Redundancy

46.1 Definitions

Nr	Dimension	Data concept	Definition	Source	Criteria not met
1	Minimum unnecessary redundancy	View	The view contains a minimum of unwanted redundancy	Redman 1996	C, K, L, M
2	Redundancy	Data	The storage of multiple copies of logically identical data	Early 2011	K
3	Redundancy	Data	The degree to which logically identical data are stored more than once.	Early 2011*	None

46.2 Notes

-

47. Relevance

47.1 Definitions

Nr	Dimension	Data concept	Definition	Source	Criteria not met
1	Relevance	Statistical information	The degree to which statistical information meets current and potential needs of the users.	Eurostat 2015	E, L
2	Relevancy	Data	The extent to which data are applicable and helpful for the task at hand.	Wang 1996	K, L
3	Relevance	Data	User-driven or application-related concern.	Redman 1996	C, H, K, L
4	Relevance	Composition of datasets	The degree to which the composition of datasets meets the needs of the data consumer.	Eurostat 2015*	

47.2 Notes

- The relevance of a dataset is dependent on its composition.
- Other dimensions that can make a dataset useful is accuracy and completeness.

48. Reliability

48.1 Definitions

Nr	Dimension	Data concept	Definition	Source	Criteria not met
1	Reliability	Initial estimated value	The closeness of the initial estimated value to the subsequent estimated value.	Eurostat, 2015	L
2	Reliability	Initial data value	The closeness of the initial data value to the subsequent data value.	Eurostat, 2015*	

48.2 Notes

- The definition of reliability implies preliminary data values and final data values.

49. Representation

49.1 Definitions

Nr	Dimension	Data concept	Definition	Source	Criteria not met
1	Representation	Data, presentation, media, documentation	Measures ease of understanding data, consistency of presentation, appropriate media choice, and availability of documentation (metadata).	CDDQ 2019	E,K

49.2 Notes

- Definition 1: See clarity.

50. Representational consistency

50.1 Definitions

Nr	Dimension	Data concept	Definition	Source	Criteria not met
1	Representational consistency	Data	The extent to which data are always presented in the same format and are compatible with previous data.	Wang 1996	C, E, K

50.2 Notes

- Definition 1: See comparability over time.
- Presentation is out of scope of this report.

51. Reproducibility

51.1 Definitions

Nr	Dimension	Data concept	Definition	Source	Criteria not met
1	Reproducibility	Dataset	The degree to which a dataset can be recreated with the same data values.		

51.2 Notes

- Reproducibility implies that the same input, the same version of the software and the same manual changes are available and can be reused.

52. Reputation

52.1 Definitions

Nr	Dimension	Data concept	Definition	Source	Criteria not met
1	Reputation	Data	The extent to which data are trusted or highly regarded in terms of their source or content.	Wang 1996	K

52.2 Notes

-

53. Robustness

53.1 Definitions

Nr	Dimension	Data concept	Definition	Source	Criteria not met
1	Robustness	View	Refers to the ability of the view to accommodate changes in the world and/or user requirements without changing the basic structure of the view.	Redman 1996	C, F, K, L, O

53.2 Notes

- The definition is not very clear. What is meant by view here?

54. Retention period

54.1 Definitions

Nr	Dimension	Data concept	Definition	Source	Criteria not met
1	Retention period	Datasets	The period that datasets are available until they can or must be deleted.		K

54.2 Notes

-

55. Scope

55.1 Definitions

Nr	Dimension	Data concept	Definition	Source	Criteria not met
1	Scope	View	The degree to which a view encompasses enough data to meet the needs of all applications and the amount of excess data.	Redman 1996	L, O

55.2 Notes

- Definition 1: See completeness (of data).
- What is meant by view?

56. Stability

56.1 Definitions

Nr	Dimension	Data concept	Definition	Source	Criteria not met
1	Stability	Information architecture or database	The degree to which information architecture or a database is able to have applications developed to use it with minimal modification of the existing objects and relationships, only adding new objects and relationships.	English 1999	E, L

56.2 Notes

- Definition 1: Concept is out of scope of the report.

57. Timeliness

57.1 Definitions

Nr	Dimension	Data concept	Definition	Source	Criteria not met
1	Timeliness	Data	Refers to several characteristics of data.	DAMA 2017	C, K, L
2	Timeliness	Data	Measure of time between when data is expected versus made available.	CDDQ	C, K, L
3	Timeliness	Data	The degree to which data represent reality from the required point in time.	DAMA-UK 2013	N
4	Timeliness	Data availability	Is the length of time between data availability and time of the event or phenomenon theydescribe.	Eurostat, 2014	K, L
5	Timeliness	Data	The extent to which the age of the data is appropriate for the task at hand.	Wang 1996	E
6	Timeliness	Data	The degree to which data is available when knowledge workers or processes require it.	English 1999	E, L
7	Timeliness	Data	The relative availability of data to support a given process within the timetable required to perform the process.	English 1999 p144	E
8	Timeliness	Data item	A measure of how current a data item is.	King 2020	K
9	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.		

57.2 Note

- Definition 2: Is related to the definition of punctuality.
- Example:

Real value creation	Data value creation	Data value available
On the date of birth	Date of registration in the birth register	Date of birth available to the data consumer.

58. Traceability

58.1 Definitions

Nr	Dimension	Data concept	Definition	Source	Criteria not met
1	Traceability	Data	The extent to which data are well documented, verifiable, and easily attributed to a source.	Wang 1996	E, K
2	Traceability	Data	The degree to which data has attributes that provide an audit trail of access to the data and of any changes made to the data in a specific context of use.	ISO 25012	E, L
3	Traceability	Data	The degree to which data lineage is available.	-	None

58.2 Note

- Traceability and data lineage are similar concepts.

59. Understandability

59.1 Definitions

Nr	Dimension	Data concept	Definition	Source	Criteria not met
1	Understandability	Data	The degree to which data has attributes that enable it to be read and interpreted by users, and are expressed in (an) appropriate languages, symbols and units in a specific context of use.	ISO 25012	E

59.2 Notes

- Definition 1: See clarity.
- Depends on availability of metadata

60. Uniqueness

60.1 Definitions

Nr	Dimension	Data concept	Definition	Source	Criteria not met
1	Uniqueness	Entity	No entity exists more than once within the dataset.	DAMA 2017	C, F, K, L
2	Uniqueness	Things	No thing will be recorded more than once based upon how that thing is identified.	DAMA-UK 2013	C, F, K, L
3	Nonduplication	Occurrences of data	The degree to which there are no redundant occurrences of data.	English 1999	E, L
4	Uniqueness	Entity	A measure of whether an entity has a single data entry relating to it within a dataset.	King 2020	K
5	Uniqueness	Objects	The degree to which objects (of the real world) occur only once as a record in a dataset.	-	None
6	Uniqueness	Records	The degree to which records occur only once in a data file.		None

60.2 Notes

Three different problems can occur:
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.

Key	Name
22	John
22	John

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.

Key	Name
22	John
37	John

c. One record has the same key as another record, and both occur in a dataset (false duplicate). Key 22 is not unique.

Key	Name
22	John
22	Peter

61. Usability

61.1 Definitions

Nr	Dimension	Data concept	Definition	Source	Criteria not met
1	Usability	Data	Is it understandable, simple, relevant, accessible, maintainable and at the right level of precision?	DAMA-UK 2013	C, F, K, L
2	Usability	Information presentation	The degree to which the information presentation is directly and efficiently usable for its purpose.	Redman 1999	G, L

61.2 Notes
- Could be used as synonym for relevance.
- Definition 1: See clarity, relevance, accessibility and granularity.
- Definition 2: Data concept is out of scope of this report.

62. Validity

62.1 Definitions

Nr	Dimension	Data concept	Definition	Source	Criteria not met
1	Validity	Data values	Refers to whether data values are consistent with a defined domain of values.	DAMA 2017	C, F, K
2	Validity	Data	Data are valid if it conforms to the syntax (format, type, range) of its definition.	DAMA-UK 2013	C, F, K, L
3	Validity	Data	The degree to which the data conforms to defined business rules.	English 1999	E, K
4	Validity	Data	Conformance of data to rules defining the syntax and structure of data.	King 2020	H, K
4	Validity	Data values	The degree to which data values comply with rules.	-	None

62.2 Notes

- A domain of values may be a set of valid values (such as in a reference table), a range of values, or a value that can be determined via rules.
- The data type, format, and precision of expected values must be accounted for in defining the domain.
- Data may also only be valid for a specific length of time, for example data that is generated from RFID or scientific datasets.
- Validate data by comparing it to domain constraints. Keep in mind that data may be valid and still not be accurate or correctly associated with particular records.

(DAMA, 2017)

- Examples of rules are metadata specifications, business rules.

63. Value

63.1 Definitions

Nr	Dimension	Data concept	Definition	Source	Criteria not met
1	Value	Data	Is there a good cost/benefit case for the data? Is it being optimally used? Does it endanger people's safety or privacy or the legal responsibilities of the enterprise? Does it support or contradict the corporate image or the corporate message?	DAMA-UK 2013	C, E, F
2	Value-added	Data	The extent to which data are beneficial and provide advantages from their use.	Wang 1996	K, E
3	Value	Data	The degree to which data provide advantages from their use.	Wang 1996*	

63.2 Notes

-

64. Variety

64.1 Definitions

Nr	Dimension	Data concept	Definition	Source	Criteria not met
1	Variety	Data	The extent to which data are available from several differing data sources.	Wang 1996	E, K
2	Variety	Data	The degree to which data are available from different data sources.	Wang 1996*	

64.2 Notes

-

65. Volatility

65.1 Definitions

Nr	Dimension	Data concept	Definition	Source	Criteria not met
1	Volatility	Data values	Remain current for a short period.	DAMA 2017	C, F, K
2	Volatility	Data values	The degree to which data values change over time.		

65.2 Notes

- In DAMA (2017), volatility is a subdimension of timeliness.
- Changes in volume of business (e.g., seasonal sales) are a type of volatility.

Version history

Version	Date	Description of the modification	Author
1.0.p1	3 March 2020	First draft based on document Andrew.	Peter
1.0.p4	25 March 2020	Draft completed.	Peter
1.0.p5	26 March 2020	Added foreword, edited text.	Andrew
1.0.p7	2 June 2020	The data concept system is taken from the report Data Concept System.	Peter
1.0.p9	3 June 2020	Edits and comments	Andrew
1.0.p10	11 June 2020	Edits and comments processed.	Peter
1.0.p11	14 June 2020	Figures adapted.	Andrew
1.0.p12	21 June 2020	Small corrections	Peter
1.0.p13	-	Lineage added. Traceability adapted.	Peter
1.0.p14	20 July 2020	Edits and comments	Andrew
1.0.p15	3 Aug 2020	Edits and comments processed. Redundancy and obtainability are added. Data file added as data concept in text and figures.	Peter
1.0	5 Aug 2020	Minor edits and comments. Final version.	Andrew
1.1	20 Aug 2020	Three data concepts added. Minor edits.	Peter
1.2	3 Sept 2020	Data categories and one source added.	Peter

Active distribution per version	
Version	**Distribution**
1.0.p1-p10	Dropbox
1.0.p11	Reviewers
1.0.p12	Chris Bradley
1.0.p13-p15	Dropbox
1.0	DAMA NL
1.1	DAMA NL
1.2	Dropbox, Website DAMA-NL