1 30 September 2022
2 Data Analytics and Methods Task Force
3 European Medicines Agency

# 4 Data Quality Framework for EU medicines regulation

5

| | |
|---|---|
| Start of public consultation | 10 October 2022 |
| End of consultation | 18 November 2022 |

6

Comments should be provided using this template. The completed comments form should be sent to dataqualityframework@ema.europa.eu

7

| | |
|---|---|
| Keywords | Data sources, studies, metadata, study protocol, study report, data flows, data management, vocabulary, glossary, use cases, population |

8



**Official address** Domenico Scarlattilaan 6 ● 1083 HS Amsterdam ● The Netherlands
**Address for visits and deliveries** Refer to www.ema.europa.eu/how-to-find-us
**Send us a question** Go to www.ema.europa.eu/contact **Telephone** +31 (0)88 781 6000 An agency of the European Union

# Table of contents

# 1. Executive Summary

As acknowledged in the recommendations of the HMA-EMA Joint Big Data Task Force (BDTF) and the workplan of the HMA-EMA Joint Big Data Steering Group (BDSG), establishing an EU framework for data quality and representativeness is a critical element for realising the full potential of (big) data and driving regulatory decisions.

This document is the first release of the EU data quality framework for medicines regulation and addresses high level principles and procedures that apply across the European Medicines Regulatory Network (EMRN)'s regulatory activities. This framework provides general considerations on data quality that are relevant for regulatory decision making, definitions for data dimensions and sub-dimensions, as well as their characterisation and related metrics. It provides an analysis of what data quality actions and metrics can be put in place in different scenarios and introduces a maturity model to drive the evolution of automation to support data-driven regulatory decision making.

This document is intended to be an overall umbrella from which more focused recommendations can be derived for specific regulatory domains with specified metrics and checks.

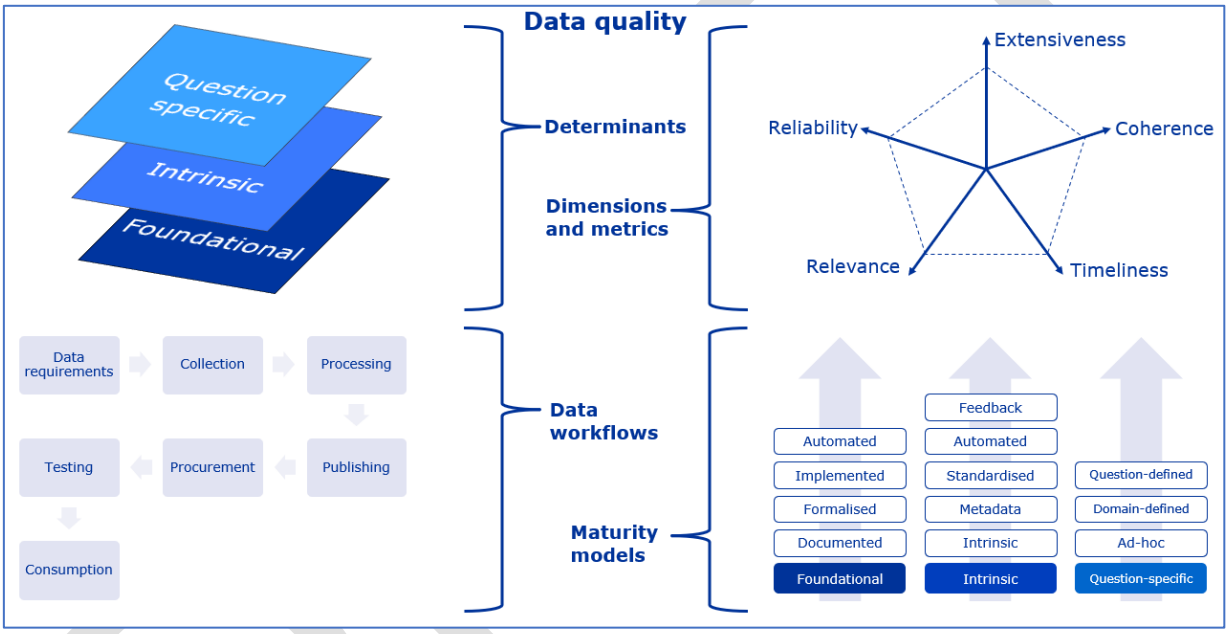


Fig 1. - Representation of the key points of the Data Quality Framework

# 2. Glossary

| | |
|---|---|
| CDM | Common Data Model |
| DQ | Data Quality |
| DQF | Data Quality Framework |
| EHR | Electronic Health Record |
| EHDS | European Health Data Space |
| EMA | European Medicines Agency |

| | |
|---|---|
| ETL | Extract, Transform and Load |
| FAIR | Findable, Accessible, Interoperable and Reusable |
| GxP | Good x Practices (where x stands for the type) - Good Laboratory Practice (GLP), Good Clinical Practice (GCP), Good Manufacturing Practice (GMP), Good Distribution/Documentation Practice (GDP) |
| ISO | International Organisation for Standardisation |
| SQuaRE | Systems and software Quality Requirements and Evaluation |
| QMS | Quality Management System |
| QSR | Quality System Regulation |

# 3. Background - the need for a Data Quality Framework for medicines regulation

As acknowledged in the recommendations of the HMA-EMA Joint Big Data Task Force (BDTF) and the workplan of the HMA-EMA Joint Big Data Steering Group (BDSG), establishing an EU framework for data quality (DQ) and representativeness is a critical element for realising the full potential of (big) data and driving regulatory decisions.

In recent years, the EU regulatory assessment process has been progressively shifting from a predominantly document-based submission to a direct assessment of the underlying data used to create those documents. This shift in process brings about the need for a framework, which would characterise the DQ and allows the regulator to make reliable assessments if the data are fit for the purpose of making decisions.

In addition, the progress in digitisation and information technology and the resulting large amount of data is creating opportunities, but also contributes to an increasingly complex landscape for regulatory decision making. As new types of data are available, guidelines or methods to demonstrate whether such data are adequate for regulatory decision making have yet to emerge. Therefore, a DQF is needed to guide coherent quality assessment procedures.

One notable example is in the increasing amount of healthcare data that are becoming available to support regulatory decision making for medicines. While clinical trials remain the fundamental method of establishing the safety and effectiveness of medicines during the pre-authorisation phase, they do not fully reflect the real world, resulting in gaps between regulatory dossiers and subsequent clinical evidence needed by downstream stakeholders including HTAs, payers and ultimately clinicians and patients. The data that European Medicines Regulatory Network (EMRN) received have the potential to bridge these gaps, but in order to realise such potential, the European Medicines Regulatory Network (EMRN) needs to acquire the ability to describe and quantify the degree to which these data are accurate and fit for purpose.

# 4. Scope of this Data Quality Framework

The scope of this DQF is to provide a set of definitions, principles and guidelines that can coherently be applied to a wide range of data sources for the purpose of characterising and assessing data quality for regulatory decision making.

As methods, terminologies, metrics and issues vary across different data types and sources this framework seeks to provide a coherent umbrella to identify, define and further develop DQ assessment procedures and recommendations for current and novel data types.

Objectives of this framework are therefore to achieve consistency in DQ related processes, enable the development of horizontal systems for DQ and eventually enable a more direct and automated use of data for regulatory decision making.

This framework builds on the recommendations of TEHDAS [1] and extends it with a classification of quality dimensions and assessment criteria, as well as with guidelines for their application. In particular, it builds on the definitions and recommendations that have been proposed in several examined DQ frameworks, including [1], [2], [3], [4], [5], [6], [7], [8], [9], [10].

## *4.1. Definition of data quality*

In general terms, quality is defined as an attribute of a product or service that defines the degree to which it meets customer and other stakeholder needs within statutory and regulatory requirements or its fitness for intended use. [2] The same principle applies to data and for the purpose of this document, we adopt the following definition:

"Data quality is defined as fitness for purpose for users' needs in relation to health research, policy making, and regulation and that the data reflect the reality, which they aim to represent" [1]

Therefore, this DQF restricts its scope to aspects of DQ that are relevant for regulatory decision making.

## *4.2. Limitations of scope*

Following the definition of DQ and the restricted focus on regulatory decision making, this framework's scope excludes:

- Evidence intended and generated insights or conclusions from underlying data. This framework focuses on defining guidelines as to assess the level of the quality of the data used for regulatory decisions, not on their actual usage for regulatory decision making and the methods involved.
- Aspects related to DQ that don't directly impact regulatory decision making e.g., conciseness, accessibility.
- Quality of the underlying elements data refers to. E.g., when considering a dataset about the purity of a medicine, this framework will cover the reliability, completeness, and other aspects of the data, but not aspects of quality (in this case purity) of the medicine per se.
- Semantic interoperability and standardisation. While these aspects are key for data usability and for the assessment of DQ, they don't relate to the assessment of quality as such. Data that are not fit for purpose in terms of answering a regulatory question won't become fit if standardised. Non-standardised data can be still theoretically used to answer a regulatory question and a DQF can also theoretically be applied to individual non-standard data sources. Therefore, the provision of guidelines and recommendations to define and select standards for interoperability shall fall out of the remits of this DQF. It falls within the scope of this document to demand the application of standards when this impacts the assessment of quality across multiple data sources.

In a similar way it is not within the scope of this guideline to provide recommendations for the specific design of systems, processes, and responsibilities to guarantee DQ, nor is it appropriate to list certain solutions or products. However, their requirement to provide evidence for DQ aspects is under scope.

This framework is intended to complement other guidelines established for the generation and management of healthcare data as to enable and optimise use in regulatory activities.

### *4.3. Structure of this DQF*

The DQF for EU medicines regulation is composed of two parts, reflecting different stages in the specification process.

The first part (general framework) is designed to provide a coherent approach to DQ that can encompass a broad range of data types and be extensible to novel use cases[1]. To achieve this, it provides a common ground on different DQ aspects that apply to different data types and scenarios: definitions, DQ dimensions and examples of metrics covering such dimensions. It furthers identifies general patterns for the applicability of DQ processes and it articulates a set of maturity models designed to drive increased automation of data-driven medicines regulatory decision making.

The second part (framework specialisation) specialises and eventually extends such generic recommendations to cater for specific data types or regulatory questions. This part poses the basis for the derivation of actual implementable guidelines, that will need to evolve as data and technologies change over time.

This document is the first version of the DQF for the European Medicines Regulatory Network (EMRN) [11], released for public consultation. It focuses on the generic framework and address the general framework, terminology, definitions, and general guiding principles around DQ in the context of medicines regulation.

In the upcoming years, the DQF will be updated on a yearly basis with further deep dives in regulatory use cases of particular interest. The document will be in line with developments in TEHDAS to further strengthen the European Medicines Regulatory Network (EMRN) data qualifications process and the collaboration with the EHDS.

## 5. General considerations underlying the maintenance and assessment of data quality

### *5.1. Data quality determinants for evidence generation*

The landscape of data that can be potentially used for regulatory purposes extends to diverse data sources, each generated through different processes and fit for different primary uses. When considering the overall quality of a dataset at the point of regulatory decision making, it is important to distinguish what contributes to quality, and what can be measured or controlled at what stage. In this framework, we classify such elements related to DQ (here referred to as "determinants" in three categories:

**Foundational determinants** pertain to the processes and systems through which data are generated, collected and made available. Foundational determinants are what affects the quality of data, but it's not part of the data themselves (and as such, they don't depend on, and cannot be derived from, the content of a dataset). For data to be trusted for regulatory decision making, we need to assess that the underlying infrastructure that collects, hosts and moves the data are designed in such a way that the correspondence between data and the real entity it represents is not altered.

[1] In the context of this framework, "use-case" is used as a broader synonym of "regulatory question", when we refer to a set of related questions and related activities.

**Intrinsic determinants** of data pertain to aspects that are inherent to a specific dataset. Intrinsic determinants are what can be derived given a dataset and possibly some external generic knowledge, but without knowledge of the context in which the data was generated, as well of the context the data will be used in (e.g., a scientific or regulatory question).

**Question specific determinants** pertain to aspects of DQ that we cannot generally define independently of a specific question.

In general, foundational determinants have a direct impact on DQ. When they cannot be controlled, the only option is to control the intrinsic aspects of DQ. The scope of such control is limited when a question (or set of typical questions) is not defined.

## *5.2. Data quality along the evidence generation process (data life cycle)*

Data that are available for evidence generation go through a process (part of a broader "lifecycle[2]") that is specific to the type of data and the larger processes and organisations that produce it.

As a reference, we can outline a general high level life cycle as follow:

- Definition of data requirements
- Data collection or generation
- Data management and processing
- Data publishing
- Data procurement and aggregation
- Testing and acceptance
- Delivery for consumption

Not all phases may be present in all data workflows (e.g., data collected from sensor or social data may be collected on a "what is available" basis, rather based on specific requirements) and possibly extra phases may apply.

For the scope of the management and assessment of DQ, it is important to assess what determinants may apply at which stage of this process, and what may be the impact. For instance, intrinsic aspects of DQ can be measured: such measures could be used to improve reliability at the stage of data collection and generation, could be used to provide an assessment of quality at publication time, must be re-assessed each time data are integrated with additional data. Question specific determinants of DQ need to be re-assessed each time data are repurposed to answer a question it was not originally collected for.

### 5.2.1. Primary vs secondary use of data

In the application of guidelines and metrics an important distinction arises between primary and secondary use of data. When systems are designed to collect and process data for a specified primary purpose, or when a set of established requirements for secondary use exist, intrinsic and question specific aspects of DQ can be already considered at the time of collection and generation. It is thus possible to design systems and processes that guarantee some quality level required for evidence generation. This is generally not the case for unforeseen secondary use of data, where the quality criteria for usage may not coincide with the ones relevant for the existing purposes of data collection. In this case, often DQ can only be controlled a-posteriori.

[2] The data life cycle is broader in that it would extend to aspects of data disposal and maintenance beyond usage.

### 5.2.2. Publication vs data consumption

Along the data life cycle, data is processed through two different contexts. In one (publication), data are generated or collected, processed, and made available. In the other (consumption), data are procured and aggregated to support analysis. These two contexts may be overlapping (e.g., when direct measurements are taken to validate a result) or may be very distinct (e.g., when data are collected and published in a catalogue for a range of possible foreseen or unforeseen usages). The overall purpose of quality assessment changes across these two contexts, and even intrinsic aspects of quality for the same dataset may differ. Detailed specification of quality assessment may be developed distinctly for these contexts, e.g., for a data catalogue, in terms of acceptable minimal quality for generic usages, or for data procurement, in terms of minimal viability for a specific question.

## *5.3. Data and Metadata*

Metadata is traditionally defined as "data about data" providing context about their purpose and generation. When data consist of numeric or unstructured information (e.g., images), metadata are typically provided as an addition to a dataset (e.g., in a file or catalogue entry). In general, the distinction between data and metadata is not well defined: some information appearing as metadata in one context (e.g., instrument provider for a test) can be considered as data in another (e.g., if assessing measurement bias).

For regulatory decision making, metadata should in general follow the same framework as data. More precisely, if some change in metadata would require a revision of the downstream generated evidence, then it should be treated as data from the perspective of DQ.

In a DQ context, metadata should not be seen as limited to metrics and summary description of datasets, but should extend to characterisation of sources, processes, and data elements definitions.

## *5.4. Frame of reference (validation vs verification)*

Some aspects of DQ can be measured in respect to different references: what is present within a dataset, or what is present beyond a dataset (this could extend to the real world). For instance, the weight of an animal could be verified for quality based on the content of a dataset (e.g., missing values), based on an overall reference or gold standard (e.g., knowledge of a natural weight range) or even verified in respect to reality.

In some frameworks, the assessment of quality within a dataset is referred to as "verification" while the assessment in respect to gold standards is referred to "validation" (this notion of validation should not be confused with validation as a form of coherence checking).

## *5.5. Granularity of data and DQ*

DQ can be assessed at different levels of granularity:

- The **value level** corresponds to a specific data point (e.g., a weight).
- The **column level** (also referred to as "variable level") covers a data point for a whole sample of individuals (e.g., weight as a variable in a clinical study DM table). Metrics for DQ at the value level are easily extended to the column level, for instance by converting binary values to a percentage.
- The **dataset level** covers an overall set of observations. In some contexts, a further distinction can be made, within a dataset, between parts of dataset that are about similar entities. When such distinction is made, we refer to such parts as "table level" (as those parts would normally appear in distinct tables).

This DQF will focus on the lowest possible level, i.e., the value level. However, some metrics will only allow the application to quality dimensions at higher level. For example, the plausibility of a single record of a person with a weight of 300 kg may not trigger a metric violation, if 80% of the records are above 300 kg it will.

# 6. DQ dimensions and metrics

The definition of DQ dimensions and metrics rely on the general definition of dimension, metrics and measures:

- A **dimension** represents one or more related aspects or features of reality (e.g., length, for a physical object).
- A **metric** represents a way to assess the value of a dimension (e.g., absolute length measured in meters in some specified circumstances).
- A **measure** represents a single data point (e.g., 2cm). More measures can be combined to derive more general metrics (e.g., average length).

DQ metrics can be defined as indicators that can be applied to a data source to derive assessments of one of more quality dimensions (a single quality metric can be used as an indicator for more than one dimension, as expressed below in the examples for coherence). For some metrics, we can define acceptance thresholds, when data is collected for a primary use case, or when some well-defined secondary uses are targeted. Such thresholds can be defined at the point of data collection. In general, and for unforeseen secondary usages, they can be defined depending on the question (or a generic set of questions) being asked.

The quality of data is the sum of several features of data, including its representation as well as its correspondence to reality. It is useful to categorise such features in dimensions, that is a set of features whose measure reveals independent aspects of DQ. In other words, different dimensions answer different distinct DQ questions.

Several data frameworks propose an organisation of DQ in dimensions, that are similar across frameworks, but often inconsistent in the exact definitions. This complicates a coherent assessment of DQ when multiple sources are aggregated. We introduce here a set of dimensions that are relevant from a regulatory point of view, complement them with a precise definition, possible metrics and examples. The intention is to remove ambiguity and provide a useful reference that can help mapping different conceptualisation of quality form a variety of sources to a common denominator that is useful to frame metrics and maturity models to support evidence generation.

## *6.1. Reliability*

We define reliability as the dimension that covers how closely the data reflect what they are designed to measure[3].

The reliability dimension answers the question: to what degree are data corresponding to reality?

When considering the "fit for purpose" definition of quality, reliability covers how correct and trustworthy the data are.

[3] This notion of Reliability is often called "accuracy" or "plausibility" in DQFs

### 6.1.1. Reliability sub-dimensions

Given our definition of reliability, we can relate other dimensions as sub-dimensions:

- **Precision** defined as the degree of approximation by which data represents reality. For instance, the age of a person could be reported in years or months.
- **Accuracy** defined as the amount of discrepancy between data and reality. This encompasses the formal definition of accuracy in measurements (e.g., the distance between the measurements and the real value) as well as measures of the amount of wrong information in a dataset. For example, the weight of a person could be given with a systematic excess weight of 1 to 2 kg if measured fully clothed.
- **Plausibility** can only be measured by confronting a data item with the entity it intends to represent and is therefore hard to measure in a data-oriented framework. Plausibility, defined as the likelihood of some information being true, is a proxy to detect errors: when some combination of information is unlikely (or impossible) to happen in the real world, this reveals accuracy issues. For example, a weight of a person exceeding 300 kg is possible, but the weight of many or all persons in a dataset exceeding that value is unplausible, likely revealing some errors in the measurement or the processing of the data.

### 6.1.2. Determinants of Reliability

Reliability fundamentally depends on the systems and process in place for the primary collection of data and its processing. In the absence of errors, accuracy would not decrease along the data aggregation process. Precision may instead decrease when data are harmonised to a common model. Intrinsic aspects of reliability are hard to measure in a pure data-oriented framework, however plausibility measures can provide a way to detect some classes of errors. Reliability is independent from a specific question, though each question, in relation to data, will set a threshold for acceptable reliability.

### 6.1.3. Reliability metrics

| Sub-dimension | Metric group | Abstract metric | Framework | Example |
|---|---|---|---|---|
| Plausibility (proxy for Accuracy) | Atemporal Plausibility | Data values and distributions agree with internal measurements or local knowledge | Validation | Height and weight are a positive value<br>Counts of unique subjects by treatment are as expected |
| | | Data values and distributions for independent measurements of the same fact agree | Verification | Oral and axillary temperatures are similar<br>Serum glucose measurement is similar to finger stick glucose measurement |
| | | Logical constraints between values agree with common knowledge | Verification | Sex values agree with sex-specific contexts (pregnancy, prostate cancer) |
| | | Values of repeated measurement of the same fact show expected variability | Verification | Weight values are similar when taken by separate nurses within the same |

| Sub-dimension | Metric group | Abstract metric | Framework | Example |
|---|---|---|---|---|
| | | | | facility using the same equipment |
| | | Data values and distributions agree with trusted reference standards | Validation | HbA1c values from hospital and national reference lab are statistically similar under the same conditions<br><br>Distribution of patients with cardiovascular disease diagnoses are similar to European Medicines Regulatory Network (EMRN) rates for the same age/sex groups |
| | | Equivalent values for identical measurements are obtained from two independent databases representing the same observations with equal credibility | Validation | Diabetes ICD-9CM and CPT codes are similar between two independent claims databases serving similar populations |
| | | Two or more dependent databases yield similar values for identical variables (e.g., database 1 abstracted from database 2) | Validation | Recorded data of birth is consistent between EHR data and registry data for the same patient |
| | Temporal Plausibility | Observed or derived values conform to expected temporal properties | Verification | Discharge date happens after admission date |
| | | Sequence of values that represent state transitions conform to expected properties | Verification | Date of an initial drug administration precedes that of the subsequent administration.<br><br>Measures of data value density against a time-oriented denominator are expected based on internal knowledge.<br><br>Count of immunisation per month shows an expected spike during flu season |
| | | Observed or derived values have similar temporal properties across one or more external comparators (gold standard) | Validation | Length of stay by outpatient procedures types conforms to insurance data for similar populations |
| | | Sequences of values that represent state transitions are similar to external | Validation | Immunisation sequences matches that of the European Medicines Regulatory |

| Sub-dimension | Metric group | Abstract metric | Framework | Example |
|---|---|---|---|---|
| | | comparators (gold standards) | | Network (EMRN) recommendations |
| | | Measures of data value density against a time-oriented denominator are expected based on external knowledge | Validation | Medications per patient-day matches claims data |

## *6.2. Extensiveness (Completeness and Coverage)*

Completeness and Coverage are two typical dimensions found in DQFs that we combine in an overarching category ("Extensiveness") as it relates to the amount of data available.

The "Extensiveness" dimension answers the question: how much data to we have? When considering the "fit for purpose" definition of quality, Extensiveness covers how sufficient are the data?

### 6.2.1. Sub-dimensions of Extensiveness

When considering the amount of information available, we can think of expressing this as a percentage respect to whole information that could be available. The distinction between completeness and coverage stems from the definition of the scope of the totally available information.

- **Completeness** measures the amount of information available with respect to the total information that could be available given the capture process and data format. Data unavailable in the dataset are called "missing". For example, the percentage of missing value for a required field (e.g., gender) in a dataset would be a measure for completeness.
- **Coverage** measures the amount of information available with respect to what exists in the real world, whether it is inside the capture process and data format or not. Coverage cannot be easily measured, as the total information may not be definable or accessible. An example of a coverage issue is whether a set of individuals present in a dataset is representative of a population under study.

A related concept to Completeness and Coverage is that of Missingness, that is meant to characterise what is the impact of incomplete data respect to coverage of a dataset.

### 6.2.2. Determinants of Extensiveness

The extensiveness of the information collected depends on the specification of the data collection process. However, when we integrate different datasets for secondary use, we have no guarantees about the completeness of the overall dataset. On a data intrinsic level, we can resort to metrics to assess the level of completeness of data. Metrics that assess how much data are present in a dataset in respect to what could be present in a given data model are simple and effective to compute. Metrics that assess how complete are the data with respect to the population they intend to measure are more complex and may involve the confrontation with gold standards. Completeness with respect to a schema is easily definable, while coverage depends on some assumptions that can be defined only at question time. At question time we will typically define a threshold (90% complete) that is acceptable for the intended question.

### 6.2.3. Metrics for Extensiveness

| Sub-dimension | Metric group | Abstract rule | Framework | Example |
|---|---|---|---|---|
| Completeness | Missing required values | Missing values respect to a local schema – over time | Verification | Breed or gender of the animal should not be NULL |
| | | Missing values respect to a local schema – single time | Verification | The encounter ID variable has missing values |
| Coverage | Estimated missing values | Missing values respect to common expectations | Verification | Lab results are missing for five consecutive days |
| | | Relative assessment of missing values respect to a trusted source of knowledge | Validation | The current encounter ID variable is missing twice as many values as the institutionally validated database<br><br>A drop in ICD-9CM codes matches implementation of ICD-10-CM |

## *6.3. Coherence*

We define coherence as the dimension that expresses how different parts of an overall datasets are consistent in their representation and meaning.

The Coherence dimension answers the question: is the dataset processable as a "whole"? Is the format of values (e.g., dates) the same across the dataset? Is the precision of values the same (e.g., age always approximated to years)? Are references to entities consistent so that information about the same entity is properly "linked" across parts of the dataset? When considering the "fit for purpose" definition of quality, coherence relates to the analysability of data.

### 6.3.1. Sub dimensions of Coherence

Coherence is a nuanced dimension which closely relates to consistency and validation. We can consider consistency and coherence largely synonyms, with the caveat that detection of inconsistencies is often a way to measure the reliability of data.

We consider the following sub-dimensions of coherence:

**Format Coherence**: whether data are expressed in the same way throughout a dataset (for instance, a data mixing dates represented as DD-MM-YYYY and MM-DD-YYYY will not be suitable for an integrated analysis).

**Structural Coherence**: whether the same entities are identified in the same way throughout a dataset. A sub-aspect of structural coherence is that references are resolved to the correct entities.

**Semantic Coherence**: whether the same value mean the same thing throughout a dataset. For instance, whether "anuria" means a condition of total cessation of urine production or the measurement of the amount of urine, or whether the same notion of a measure is intended to have the same precision throughout a dataset.

**Uniqueness**: for the scope of this framework, we consider uniqueness as sub-dimension of coherence. Uniqueness is the property that the same information is not duplicated but appears in the dataset once. This problem is typical for linked data from different sources.

Strictly related to coherence are **Conformance** and **Validity**.

**Conformance** relates to coherence in that it assesses coherence toward a specific reference or data model. Conformance may practically be the best way to assess coherence, and it also specialised as format, structural and semantic conformance. **Validity** is a narrower case of conformance that is defined when the reference model is specific to the dataset being assessed.

### 6.3.2. Determinants for Coherence

Coherence of data at source largely depends on the synchronisation of processes and systems across an organisation generating data or, when multiple data are aggregated, on the commitment of such organisation to the use of internal or external data standards. By extension, coherence for data aggregated and repurposed for secondary usage depends on the availability of shared standards and reference data. The intrinsic nature of the coherence of a dataset can be improved, largely within a data processing steps. However, when improving coherence involves approximating or clarifying the meaning of data, access to the source system and processes is often required (e.g., for clarifications). Some aspects of semantic coherence may be difficult to assess with a metric and hence only comparable at query time.

### 6.3.3. Metrics for Coherence

| Sub-dimension | Metric group | Abstract rule | Framework | Example |
|---|---|---|---|---|
| Format coherence (conformance) | Syntactic constraints | Data Values conform to internal formatting constraints | Verification | Sex is only one ASCII character |
| | Allowed values | Data values conform to allowable values or ranges | Verifcation | Sex for the animal only has values "M", "F". or "U" |
| | | Data values conform to the representational constraints based on external standards | Validation | Values for primary language conform to ISO standards |
| Relational coherence (conformance) | Reference coherence | Data values conform to relational constraints | Verification | Patient medical record number links to other tables as expected |
| | | Unique (key) data values are not duplicated | Verification | A medical record number is assigned to a single patient |
| | | Data values conform to relational constraints based on external standards | Validation | Data values conform to all not-NULL requirements in a common multi-institutional data exchange format |
| | Schema coherence | Changes to the data model or data model versioning | Verification | Version 1 data does not include medical discharge hour |

| Sub-dimension | Metric group | Abstract rule | Framework | Example |
|---|---|---|---|---|
| | Computational coherence | Computed values conform to programming specifications | Verification | Database calculated and hard calculated BMI (body mass index) values are identical |
| | | Computed results based on published algorithms yield values that match validation values provided by external sources | Validation | Computed BMI percentiles yield identical values compared to test results and values provided by the European Medicines Regulatory Network (EMRN) |
| Semantic coherence (conformance) | Precision coherence | The precision of values is fitting a target standard | Verification | E.g., two decimal digits are used and generally not zero. |
| | Semantic coherence | Use of code lists is consistent across data | Verification | E.g., the level of a MedDRA coding for an indication doesn't vary across the dataset. |
| Uniqueness | | Same subject is represented with the same identity | Verification | William Smith is also represented as Bill Smith with the same DOB |
| | | Same subject is represented with multiple identities | Verification | William Smith and William Smith appear as separate individuals instead of the same individual |
| | | The data records of individuals are matched using unique keys | Validation | William Smith's DOB ID matches with Bill Smith's DOB and ID |

## *6.4. Timeliness*

We define timeliness as the availability of data at the right time for regulatory decision making, that in turns entails that data are collected and made available within an acceptable time.

The timeliness dimension answers the question: is this data reflecting the reality at the desired point of time?

When considering the "fit for purpose" definition of quality, timeliness covers how closely the data reflects the reality at the time it intends to measure.

### 6.4.1. Sub-dimensions of Timeliness

**Currency** is a specific aspect of timelines that considers how fresh is the data (e.g., current and immediately useful).

In the context of our framework **lateness**, intended as the aspect of data being captured later than expected, falls in the dimension of reliability (is this data corresponding to reality?).

### 6.4.2. Determinants of Timeliness

Timeliness is determined by the systems and processes used to collect and make data available.

### 6.4.3. Metrics for Timeliness

| Sub-dimension | Metric group | Abstract rule | Framework |
|---|---|---|---|
| Currency | | The average time of updates in a database (or timestamp) | Verification |
| | | The last update of a database (or timestamp) | Verification |

## *6.5. Relevance*

Relevance is defined to the extent to which a dataset presents data elements useful to answer a research question. While a broad notion of relevance encompasses all aspect of quality, we focus here on the narrower aspect of what data elements are present.

The relevance dimension answers the question: does the dataset present the kind of values that we need to address a specific question?

When considering the "fit for purpose" definition of quality, relevance covers how closely the data reflects the aspects of reality that we intend to measure.

### 6.5.1. Determinants of Relevance

Relevance can only be established in relation to a regulatory question. In some cases, it is possible to identify a set of typical questions that cover the need of a coherent range of usages for some data types. We can then establish relevance with respect to such questions, or in short **relevance for a domain**.

### 6.5.2. Metrics for Relevance

| Sub-dimension | Metric group | Abstract rule | Framework | Example |
|---|---|---|---|---|
| | | The fraction of required variables (columns) available in a given dataset | Verification | |

# 7. General recommendations and maturity models

Selecting data assets to use in regulatory decision making ultimately requires knowledge of the degree to which such asset satisfies reliability, extensiveness, coverage, coherence, and relevance criteria. Such quality dimensions build up along an overall life cycle from generation to processing to aggregation and ultimately analysis, and in such process, data originally gathered for other usages can be repurposed (when ethical or legal requirements are met [12]).

The choice of quality measures and checks varies broadly depending on data types and their intended use. However, it is possible to organise such measures and checks following a coherent structure, that help achieving homogeneity and identify gaps.

The following tables exemplify how determinants of quality (foundational, intrinsic or question specific) affect the different quality dimensions and how, for both data and metadata. These tables provide a guidance for what metrics and actions apply at which stage of the data lifecycle. For example, the

dimension of extensiveness is determined exclusively by foundational determinants (e.g., at production time). Further in the data life cycle, data intrinsic measures can only partially assess the degree of reliability (plausibility metrics).

These tables also form the basis for the development of maturity models for the characterisation of DQ for regulatory purposes. The maturity models provide guidance as to how determinants can be characterised, in successive level of maturity, that increase by the progress toward the strongest possible evidence generated in the most efficient way to support regulatory decision making.

| Determinant/ Dimension | Reliability | Extensiveness | Coherence | Timeliness | Relevance |
|---|---|---|---|---|---|
| **Foundational** | **Primary and secondary**<br>Data reliability results from systems and processes in place for data generation or collection.<br>Reliability is affected by data processing and transformations at later stages<br>**Secondary**<br>Precision may decrease during data transformation harmonisation processes | **Primary and secondary**<br>The data collection protocol determines what data are collected.<br>**Primary**<br>Data collected following established protocols can be sufficient to address regulatory questions.<br>**Secondary**<br>There is no guarantee on the completeness of an integrated dataset or its coverage for a different use case, and this can only be assessed or controlled. | **Primary and secondary**<br>Dependent on the orchestration of processes originating data and on the commitment to internal or external data standards.<br>**Secondary**<br>Relies on shared standards and reference data.<br>Documentation on data generation processes may be needed to enhance coherence. | **Primary and secondary**<br>Solely determined by systems and processes. | **Primary**<br>Normally guaranteed by the design of the data collection process.<br>**Secondary**<br>Normally assessed for a specific use or a class of usages when datasets are selected. |
| **Intrinsic** | **Primary and secondary**<br>Plausibility measures can be used to detect a (limited) class reliability issues.<br>Direct measures of accuracy require | **Primary and secondary**<br>Completeness measures based on a data model are easy to implement.<br>**Secondary**<br>Coverage measures are more complex and may require confrontation to | **Primary and secondary**<br>Coherence can be measured exclusively based on data (with eventual access to datasets-independent reference data).<br>**Secondary**<br>Coherence can be largely | **Primary and secondary**<br>Some aspects of timeliness may be observed in the datasets (e.g., event dates to determine currency).<br>A dataset itself cannot in general reveal how current its information is. | **Primary and secondary**<br>Relevance of data are not dependent on a dataset itself. |

| | | | | | |
|---|---|---|---|---|---|
| | access to the source of data. | a golden standard. | improved based solely on a dataset and data-independent elements (e.g., mapping to a common standard).<br><br>A full resolution of coherence may require access to additional information on processes.<br><br>Coherence needs to be assessed every time a new data source is "integrated". | | |
| **Question specific** | **Primary**<br><br>Processes and systems to collect data are usually designed to answer a specific question and to meet the required targets, across DQ dimensions, that such target entails. | | | | |
| | **Secondary**<br><br>Threshold for acceptable reliability can be defined at question time. | **Secondary**<br><br>Coverage and completeness depend on a question: metrics can be defined at question time or for a domain.<br><br>For completeness, typically a question would determine a set of acceptance thresholds and general metrics. | **Secondary**<br><br>Some assessment of semantic coherence (data distribution coherence or abstraction coherence) may only be measured at question time. | **Secondary**<br><br>Acceptable timeliness depends on the question and its broader regulatory usage (e.g., approval vs monitoring). | **Secondary**<br><br>Relevance can only be determined in relation to one or more questions. |

459

460 Determinant to quality dimension implications, Data.

| Determinant /Dimension | Reliability | Extensiveness | Coherence | Timeliness | Relevance |
|---|---|---|---|---|---|
| **Foundational** | **Primary and secondary**<br><br>Reliability of Metadata relies on the processes to collect it, along the | **Primary**<br><br>For primary data, the extensiveness of metadata can be | **Primary**<br><br>Metadata coherence relies on the presence of common | **Primary and secondary**<br><br>Timeliness of Metadata are purely dependent on the processes | **Primary**<br><br>Normally guaranteed by the design of data collection process. |

| | | | | | |
|---|---|---|---|---|---|
| | whole data processing chain.<br><br>One key aspect to ensure reliability is to capture metadata as close to the source as possible. | characterised at source. | standards and terminologies.<br><br>**Secondary**<br><br>For secondary data, coherence relies on the presence on widely agreed standards and shared resources such as ontology or reference data services. | supporting its collection.<br><br>**Secondary**<br><br>When data are repurposed and used in different systems, timeliness of metadata should be enforced by design (metadata should be in synch with the data) | **Secondary**<br>Relevant metadata can be required and controlled by a downstream system but cannot be guaranteed at source. |
| **Intrinsic** | **Primary and secondary**<br><br>Some metadata (e.g., summary statistics) can be generated from a dataset | **Primary and secondary**<br><br>Intrinsic measures for meta DQ mimic the ones for data (e.g., completeness and missing fields).<br><br>Unlike data, metadata assessment may not require references to golden standards (e.g., missing metadata values is not related to sampling of a population) | **Primary and secondary**<br><br>Metadata coherence solely depends on a specific metadata and data-independent elements (e.g., shared reference data). | **Primary and secondary**<br><br>The assessment of timelines aspect of data typically depends on metadata (e.g., timestamps) | **Primary and secondary**<br><br>Relevance of metadata does not depend on a dataset itself. |
| **Question specific** | **Primary**<br><br>Metadata requirements are designed for a specific question and are normally sufficient to address it. | | | | |
| | **Primary and secondary**<br><br>Metadata should be in general reliable independentl y of a specific question (not all metadata collected may be relevant for | **Secondary**<br><br>The characterisation of what metadata are necessary is ultimately dependent on a question (or set of typical questions) | **Primary and secondary**<br><br>The coherence of metadata is independent from a specific question. | **Primary and secondary**<br><br>Timeliness of metadata are independent from a specific question. | **Secondary**<br><br>Relevance of metadata are purely dependent on a question (or range of questions). |

| | | | | | |
|---|---|---|---|---|---|
| | all questions). | | | | |

Determinant to quality dimension implications, Metadata.

## *7.1. Foundational determinants: Recommendation and maturity levels*

A characterisation of the systems and process underpinning data generation and processing (foundational determinants) is necessary to assess DQ. We provide here a set of maturity levels, each providing a progressive set of recommendations for the characterisation of foundational determinants, with the intention to chart a direction of improvement towards an increased, supported by large scale, assessment of evidence.

### 7.1.1. Level 1: documented

For data to be used in regulatory decision making, at a minimum, the processes that pertain to data generation and manipulation should be **documented**, true and **verifiable** (when relevant, this may extend to training procedures). This is fundamental to ensure the reliability of any derived information and documentation and should cover determinants for reliability (precision), extensiveness, coherence and (when relevant), timeliness (while some of these depends on a specific question, data collection processes and systems will generally be designed with some primary questions as a reference). The provision of documentation for data processing and transformation are also essential to guarantee that reliability is preserved and should be provided for all such processing by different actors along the data life cycle.

From a metadata perspective, this means metadata (in some form) should always accompany a dataset it refers to.

In order to guarantee the truth (correctness of data) **audit** procedures or other controls should be in place.

When a system is designed for continuous data collection (as opposed to a one off), additional processes of **performance monitoring** and improvement should be in place.

### 7.1.2. Level 2: formalised

The second level of the maturity model includes and extends the first level, by requiring that, whenever possible, documentation and metadata should be following an industry standard framework. Level 2 should be considered the minimal level of acceptable maturity, though exceptions may arise for novel data types. The recommendation to use standards extends to metadata.

### 7.1.3. Level 3: implemented

Systems are in place that implement industry standard DQ processes automatically and by design. A range of infrastructure should be in place to support data management, including support for standardisation (e.g., reference data management). By reducing the scope for human errors, such an implementation can generally improve reliability and coherence (e.g., respect to multiple interacting processes). Such an implementation may also be necessary to guarantee timeliness and it should ensure that metadata are collected by design, and as close to the data generation events as possible.

### 7.1.4. Level 4: automated

The operations and output of the above systems and infrastructure should be machine readable, as to unify data and DQ elements for direct downstream consumption. Metadata should be represented following FAIR principles. This is intended to be an aspirational level.

## *7.2. Intrinsic determinants: Recommendations and maturity levels*

Beyond documented evidence of how data was collected or generated, we can typically apply measures of intrinsic aspects of DQ. These can be directly derived from the dataset, but their computation could also rely on some external body of knowledge.

### 7.2.1. Level 0: intrinsic

There are no hard minimal requirements for quality, as any piece of evidence can be assessed before being used to generate evidence. Nevertheless, the propagation of data without an associated quality assessment should be discouraged.

### 7.2.2. Level 1: metadata

Data are provided with a set of quality metrics as metadata. Some of these data can be directly derived from the dataset, while other derive from the overall data collection process (e.g., sampling, bias). Metadata should also cover the description of data elements that are necessary for its interpretation.

### 7.2.3. Level 2: standardised

Data are provided with a standardised set of quality metrics, that can be compared across datasets. When applicable or possible, standards should extend to cover reference knowledge that can be used to assess a dataset in respect to what is meant to represent (e.g., typical population distributions to assess biases). Metadata makes use of shared definitions, that also enable comparability and integration across datasets.

### 7.2.4. Level 3: automated

Quality assessment is automated (at least for a large extent of metrics). In general, this is feasible only when data are expressed in a common data model, so that a standard library of tests can be run on incoming data. Data and metadata should follow FAIR principles.

### 7.2.5. Level 4: feedback

There is a data ecosystem in place so that quality assessment by data consumers can provide feedback to improve the data collection and production process.

(Note that the order of maturity of level 2 and 3 may change for particular data types.)

## *7.3. Recommendations and maturity levels for question-specific aspects of data quality*

In general, it is not possible to assess the relevance of a dataset, as well as aspects of extensiveness and precision, without a target question. However, when considering the adoption of a large body of

data for regulatory decision making, and its possible use beyond primary use cases, it becomes important to articulate to what degree DQ, including relevance, can be assessed "a-priori".

### 7.3.1. Level 1: ad-hoc

All dimensions that are question specific are assessed only at "query time" on an ad hoc basis.

### 7.3.2. Level 2: domain-defined

A range of common questions is identified, from which metrics and thresholds can be derived that can be used to guarantee acceptable levels of quality. Data published in data catalogues should make use of such metrics.

### 7.3.3. Level 3: question-defined

The requirements for a specific question are precisely codified and can be mapped to metrics and thresholds in a way that could automatically assess the relevance of a dataset for a specific question. This is the natural level for primary use cases, while for secondary use of data this should be intended as an aspirational level.

## *7.4. Quality at source*

As a general guideline, in designing data collection and generation processes, aspects of DQ should be addressed as early as possible. For instance, assessment of quality done close to the moment of production can help correcting a collection error. The further data travels from the original context, the harder it becomes to correct issues. This is particularly relevant for metadata as knowledge of the context of data generation is maximally present only at generation time.

## *7.5. The role of QMS*

A Quality Management System (QMS) [1,3] is a formalised approach adopted by an organisation that documents processes, procedures, and responsibilities for achieving quality policies and objectives. It achieves these quality objectives through quality planning, quality assurance, quality control and quality improvement. Whenever possible DQ processes should be framed in the context of standard QMS. In particular, standards like the ISO 9000 family define QMS across industries, while more specific QMS have been developed for specific industry or products (e.g., ISO 2500 for software products).

# 8. Regulatory use of data for decision making

The generic framework here introduced is intended to be applied to a wide range of regulatory decision making based on evidence generated through data analysis in the context of medicinal products evaluation and monitoring. Among these areas, a few have been identified as areas of special in relation to this DQF: bioanalytical omics data, animal health data, preclinical data (cell-based and animal-based laboratory data), spontaneous adverse drug reporting data, chemical and manufacturing control data.

# 9. References


[1] European Health Data Space Data Quality Framework, Deliverable 6.1 of TEHDAS EU 3rd Health Program (GA: 101035467). May 18th, 2022, accessed at https://tehdas.eu/results/tehdas-develops-data-quality-recommendations

[2] ISO 9001:2015 Quality Management System, accessed at https://www.iso.org/standard/62085.html

[3] Kahn MG, et al. A Harmonized Data Quality Assessment Terminology and Framework for the Secondary Use of Electronic Health Record Data. EGEMS (Wash DC). 2016 Sep 11;4(1):1244. doi: 10.13063/2327-9214.1244.

[4] Healthcare Data Quality: A 4-Level Actionable Framework [Internet]. 2020 [cited 2022 Sept 5th]. Available from: https://www.healthcatalyst.com/insights/healthcare-data-quality-4-level-actionable-framework

[5] Schmidt CO, Struckmann S, Enzenbach C, Reineke A, Stausberg J, Damerow S, et al. BMC Framework - Facilitating harmonized data quality assessments. A Data Quality Framework for observational health research data collections with software implementations in R. BMC Medical Research Methodology. 2021 Apr 2;21(1):63.

[6] Sentinel QA Program. Quality Assurance - Sentinel Version Control System [Internet]. [cited 2022 Feb 14]. Available from: https://dev.sentinelsystem.org/projects/QA/repos/qa_package/browse

[7] NESTcc. Data Quality Framework, A report of the Data Quality Subcommittee of the NEST Coordinating Center - An initiative of MDIC [Internet]. 2020 [cited 2022 Feb 14]. Available from: https://nestcc.org/nestcc-data-quality-framework

[8] The National Patient-Centered Clinical Research Network. PCORnet - Data Quality Framework [Internet]. [cited 2022 Feb 18]. Available from: https://pcornet.org/data

[9] Duke-Margolis Center for Health Policy. Characterizing RWD Quality and Relevancy for Regulatory Purposes [Internet]. 2018. Available from: https://healthpolicy.duke.edu/sites/default/files/2020-03/characterizing_rwd.pdf

[10] Duke-Margolis Center for Health Policy. Determining Real-World Data's Fitness for Use and the Role of Reliability [Internet]. 2019. Available from: https://healthpolicy.duke.edu/sites/default/files/2019-11/rwd_reliability.pdf

[11] Big Data Steering Group, Big Data Workplan 2022-2025. [Internet] 2022. Available from: https://www.ema.europa.eu/en/news/big-data-use-public-health-publication-big-data-steering-group-workplan-2022-25

[12] European Commission. Can we use data for another purpose? [Internet, cited 27 Sept 2022] https://ec.europa.eu/info/law/law-topic/data-protection/reform/rules-business-and-organisations/principles-gdpr/purpose-data-processing/can-we-use-data-another-purpose_en[12] American Society for Quality. What is a Quality Management System (QMS)? | ASQ [Internet]. [cited 2022 Feb 18]. Available from: https://asq.org/quality-resources/quality-management-system