



EUROPEAN MEDICINES AGENCY
SCIENCE MEDICINES HEALTH

Session 1: Data Quality Framework overview

EU Data Quality Framework and development of the ADR deep-dive chapter

Presented by Ana Cochino (EMA)

Multi-stakeholder workshop on data quality framework for Adverse Drug Reaction reporting– 01 03 2024

An agency of the European Union



Topics outline

- 1 Background and scope of the EU-DQF activities
- 2 Timeline – activities completed and next steps
- 3 Dimensions and sub-dimension – measuring Data Quality
- 4 Data Quality maturity models
- 5 Application of the DQF in the ADR context – points to consider

- The drafting work started under BDSG (Big Data Steering Group) recommendation (HMA-EMA joint working group) on **developing a Data Quality Framework**
- After the **set-up of the MWP** (Methodological Working Party) **of the CHMP**, the sponsorship of this deliverable has been co-shared between the two groups – published as **CHMP approved guideline**
- The document has been developed in close **collaboration** with experts from various regulatory fields (e.g.: quality, pre-clinical, ADR, veterinary) and widely consulted
- EU DQF is describing concepts **applicable to all data sources used in medicine regulation** across EU regulatory network
- **Further deep-dives** (= applications, specific use cases) will be developed in the following iterations

DARWIN EU
Data quality & representativeness
Data discoverability
EU Network skills
EU Network processes
Network capability to analyse
Delivery of expert advice
Governance framework
International initiatives
Stakeholder engagement
Veterinary recommendations

2022

EU Data Quality Framework

- Landscape analysis of existing data quality frameworks and similar initiatives
- Setting up a workgroup with representatives of several regulatory use-cases (e.g.: veterinary domain, pre-clinical, ADR)
- April 2022: EU-DQF multi-stakeholder workshop
- November 2022: publication of the data quality framework draft for public consultation

2023

EU Data Quality Framework

- Addressing of comments received during public consultation
- June 2023: second round of CHMP consultation and re-drafting
- December 2023: publication of a finalised version of the [EU Data Quality Framework](#), adopted by CHMP

DQF-RW

- June 2023: multi-stakeholder workshop on Data Quality Framework in the context of Real-World Data Sources
- A first draft of the DQF-RW was produced and underwent internal revision

DQF – ADR

- Preparatory work for ADR-DQF
- Preparatory work for a multi-stakeholder workshop

2024

- Preparatory work for ADR-DQF (ongoing)
- March 2024: Multi-Stakeholder workshop: Data Quality Framework for ADR Data

Next steps 2024:

EU Data Quality Framework

- Prioritisation of use-cases to be addressed going forward

DQF-RW

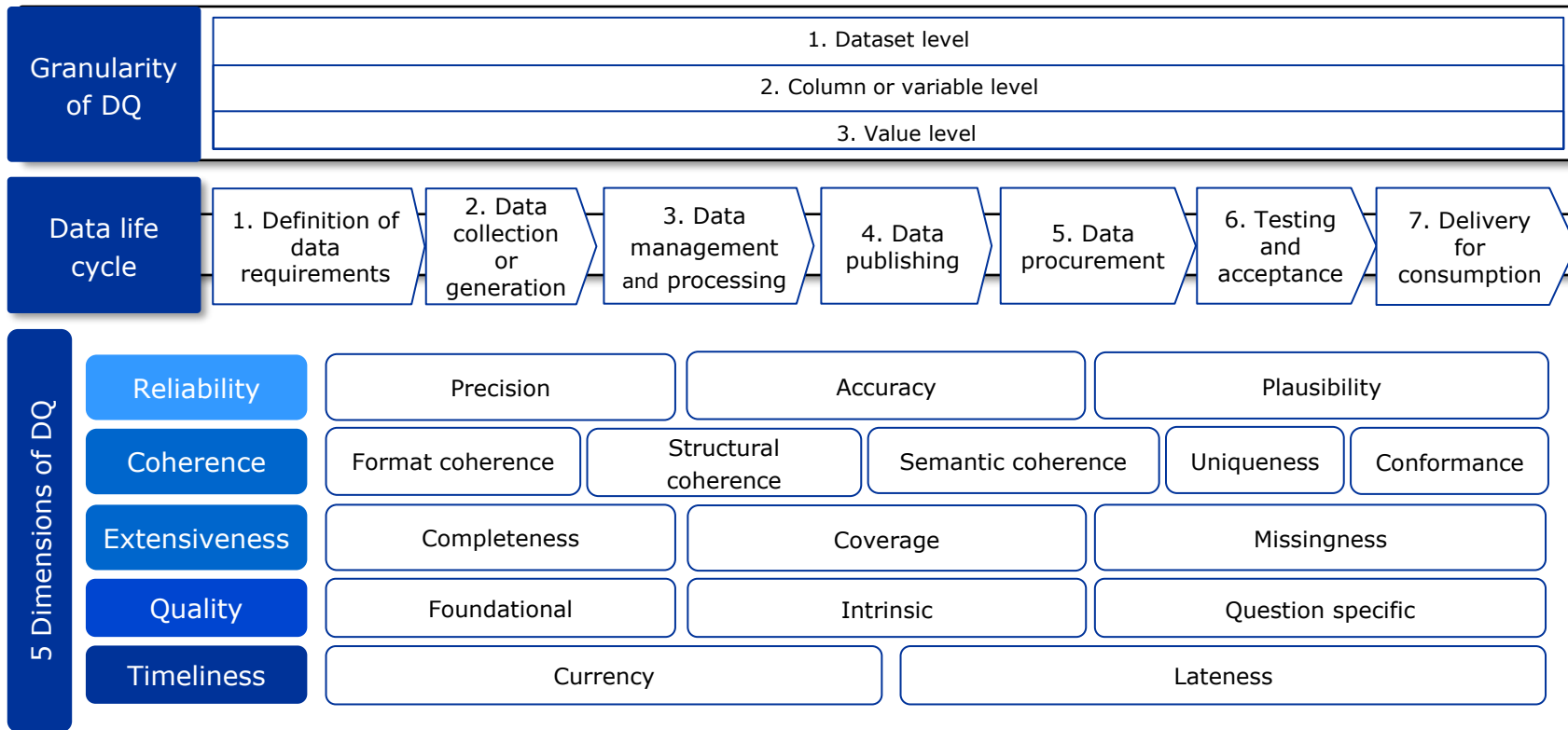
- Regulatory Network consultation of the DQF-RW draft
- Re-drafting based on comments
- Public consultation of the DQF-RW and aiming for publication (tbc)

DQF - ADR

- Preparing of a first draft of ADR-DQF
- Consultation process to be defined (timelines and sequence of steps)

- The EU Data Quality Framework guideline seeks to address the topic of characterising the data quality of any given dataset used for medicine regulation
- It sets out the concepts, the definitions, the metrics and proposes a set of maturity models to describe the quality of a dataset
- Its main focus is measuring, describing and where possible define standards for what can be considered a 'fit for purpose' dataset
- Aims to provide a coherent basis to identify, define, and further develop DQ assessment procedures and recommendations for current and novel data types





In the EU DQF, we partition data quality aspects ("determinants") in three different categories



What pertains to quality that cannot be assessed or measured without a defined research question and method

What quality aspects can be assessed or stated about a dataset, independently from the way this was generated

What contributes to the quality of a dataset, that is not part of a dataset itself

Main dimensions of quality as outlined in the general Data Quality framework



Is the data correct?
Is it representing what is meant to represent?

RELIABILITY



How much data is there?

EXTENSIVENESS



Can data be analysed as a whole?

COHERENCE



Is data available at the right time?

TIMELINESS



Is this the kind of data I need?

RELEVANCE*


To what degree are data corresponding to reality?



- **Precision** defined as the **degree of approximation** by which data represents reality (e.g.: the age of a person could be reported in years or months).
- **Accuracy** defined as the **amount of discrepancy between data and reality**. (e.g.: errors in recording the weight by mistype or systematic by measuring it fully-clothed)
- **Plausibility** defined as the likelihood of some information being true. (e.g.: a weight of more than 300 kg is *likely* an error)



How much data do we have"? When considering the "fit for purpose" definition of quality, extensiveness covers how sufficient the data are



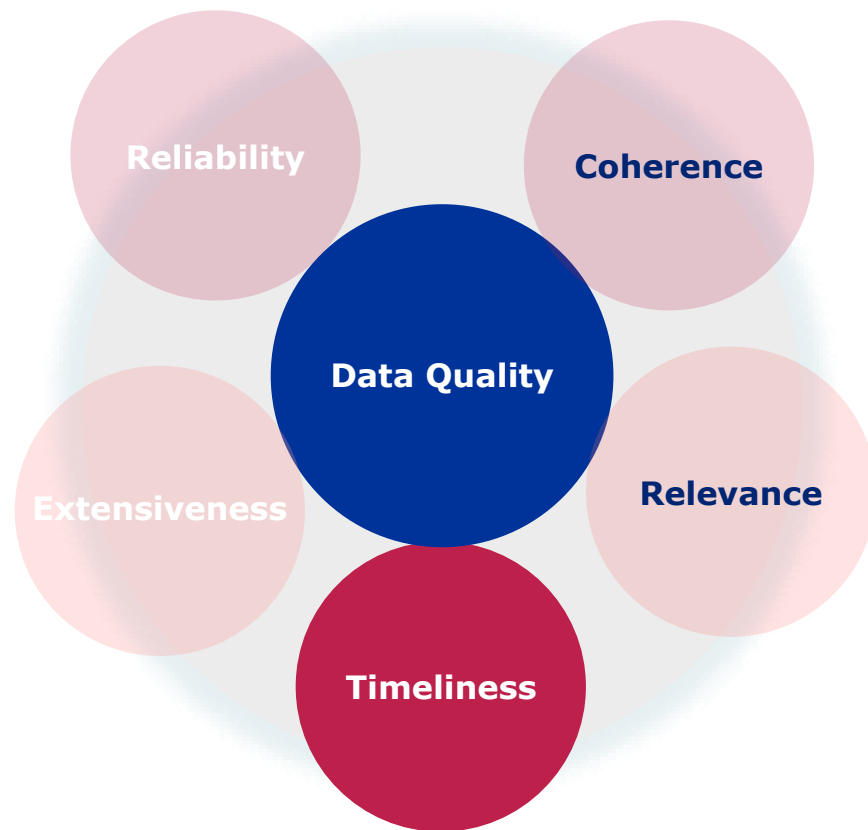
- **Completeness** defined as the amount of information **available** with respect to the total information that **could be available** given the capture process and data format (e.g.: % of non-missing values for gender in a dataset).
- **Coverage** defined as the amount of information **available** with respect to **what exists in the real-world**, whether it is inside the capture process and data format or not (e.g.: % of the population of a given country available in a dataset)



Timeliness is defined as the availability of data at the right time for regulatory decision making



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



Relevance is defined to the extent to which a dataset presents data elements useful to answer a research question



Does the dataset present the kind of values that we need to address a specific question, using a specific method?



Coherence expresses how different parts of an overall dataset are consistent in their representation and meaning



- **Format coherence** : data **expressed** in the same way throughout the dataset
- **Structural coherence**: entities **identified** in the same way throughout the dataset
- **Semantical coherence**: the same value have the same **meaning** throughout the dataset
- **Uniqueness**: same information is **not duplicated**, but appears in a dataset once.



Characterisation of determinants for the development of maturity models

Foundational

Documented

Formalised

Implemented

Automated

Intrinsic

Intrinsic

Metadata

Standardised

Automated

Feedback

Question-specific

Ad-hoc

Domain-defined

Question-defined

- The applicability to non-EU data sources and alignment with other DQFs and opportunities for international convergence should be considered
- The **heterogeneity of different EU systems** needs to be taken into account
- Guidance or tools to help implement the maturity model for different regulatory use cases would be beneficial (e.g., checklists, score cards, details about required documentation for quality reports)
- Implications for **system owners** and **data submitters** should be well articulated
- Precise measurements of data quality might not be always necessary; in relation with some uses, it might be sufficient to **request descriptive documentation** on data reliability and relevance (as opposed to validation/verification)
- **Primary use of data** (data used for the initially designed purpose) is characterised by a better control of the data quality management; consider also the **secondary** use of ADR data (e.g.: in the context of RWD or for additional analysis)

- EU Data Quality Framework – guideline [published](#) after CHMP adoption, applicable to all data used for regulatory purpose
- Concepts defined for the overarching use case in medicine regulation are being further refined and adapted to individual use cases (deep-dive)
- The drafting process involves wide stakeholder consultation and communication
- Questions, comments, proposals - reach out to DataQualityFramework@ema.europa.eu