EMA
EUROPEAN MEDICINES AGENCY

ADR Data Quality Framework

Development, summary and next steps

Tom Paternoster-Howe, EMA

- Background
- Status
- Challenges – Dimensions
- Recommendations
    1. Reliability
    2. Extensiveness
    3. Coherence
    4. Timeliness
    5. Relevance
    6. Special considerations for specific ADR data sources
- Conclusion & next steps

Background

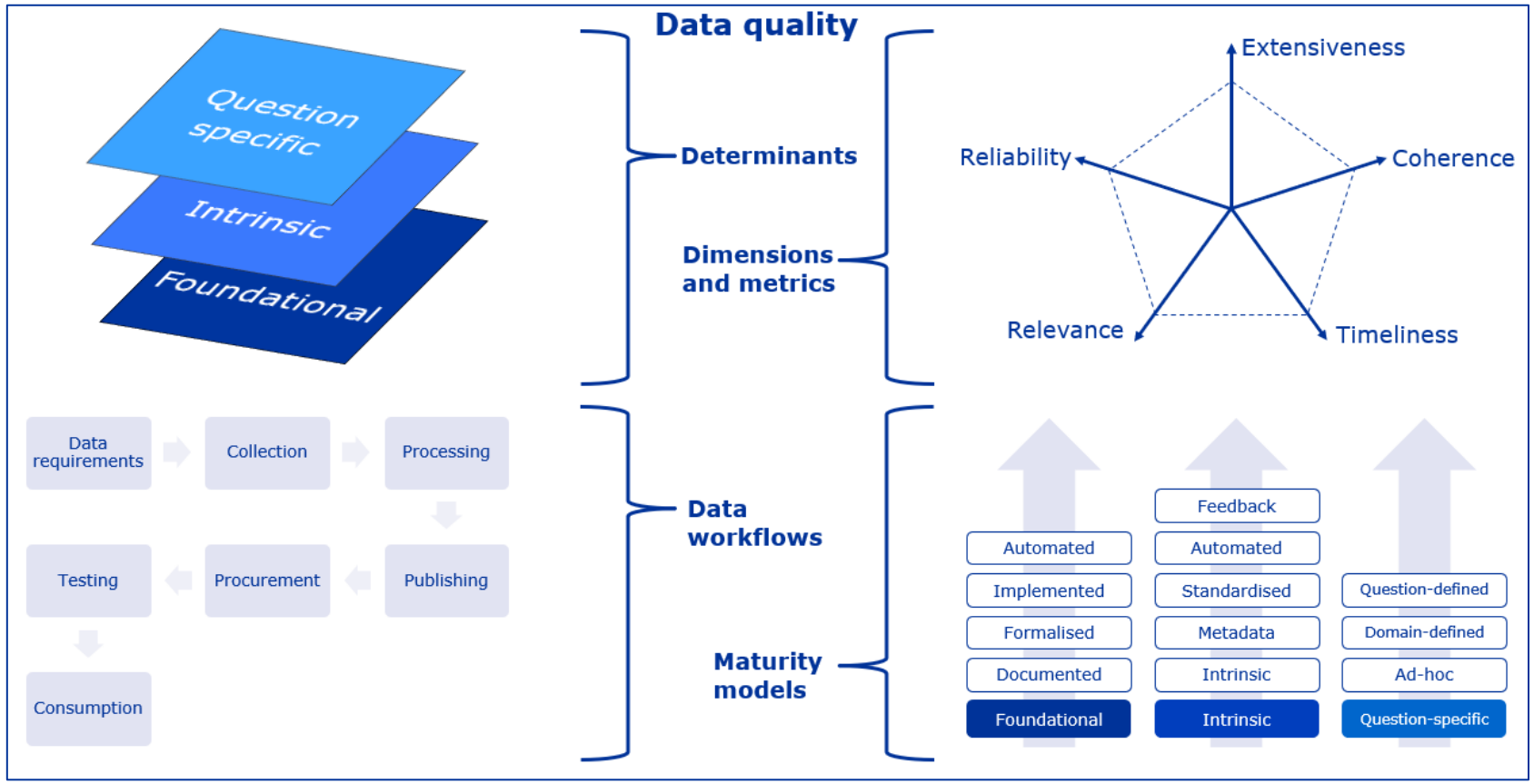
EU Data Quality Framework (EU DQF) – background

- The EU Data Quality Framework guideline has been developed under BDSG (Big Data Steering Group) recommendation (HMA-EMA joint working group), MWP (Methodology Working Party) of the CHMP
- CHMP adopted the document in October 2023, following a public consultation
- EU DQF is describing concepts and metrics applicable to all data sources used in medicine regulation across EU regulatory network.
- It defines the terminology around data quality, providing the regulator with a framework to measure and describe the data quality of a given data source, such as:
  - Data quality aspects of the data lifecycle (‘foundational determinants’) – a structured approach to data quality from the perspective of data generation and its processes
  - Looking at dataset and its characteristics, with a set of metrics and dimensions to measure and assess the ‘intrinsic determinant’
  - Aspects of data quality from the perspective of its use (e.g.: a safety concern that needs further investigation, or the generation of a signal) – ‘question specific determinants’

Fundamental concepts described (for illustration)

Data Quality can be characterised using *different concepts*

- Determinants
- Dimensions and metrics
- Data workflows
- Maturity models



EU Data Quality Framework (EU DQF) – applications

Further deep-dives (= applications, specific use cases) are planned to be developed in the following iterations of the DQF work, aiming to:

- Apply the concepts described specifically to medicine regulation domains (e.g.: medicinal product information, adverse drug reaction data, real-world data)
- Describe metrics and data quality maturity models that will characterise the data quality, specifically applied to the given domain
- Identify any potential gaps in data quality assessment and address them
- Draft a dedicated guideline
    - With a separate chapter for each medicine regulation domain

Application to ADR data

- EMRN organised a multi-stakeholder workshop to obtain input from all stakeholders involved in ADR reporting
    - Stakeholder groups represented included multinational MAHs, SME MAHs, Generic MAHs, Commercial Sponsors, Non-commercial Sponsors, Investigators, Ethics Committees, NCAs (PV), NCAs (CT), PV Inspectors, veterinary MAHs and veterinary NCAs
- The aim of the workshop was to bring together experts in the field to build on their extensive experience and knowledge relating to ADR data quality, in order to:
    - Familiarise key stakeholders with the published EU Data Quality Framework (DQF)
    - Collect input from experts in the field and learn from existing experiences
    - Discuss important challenges related to measuring, characterising and improving data quality in the context of Adverse Drug Reaction (ADR) data.
    - Identify the key topics which should be covered in the Data Quality Framework ADR chapter

Status

Status of DQF ADR chapter

- Multi-stakeholder workshop ✓
- One-to-one interviews with various stakeholders ✓
- First draft to EMA ✓
- Feedback from EMA ✓
- Final draft to EMA ✓
- Consult with regulatory stakeholders regarding recommendations ✓
- Prepare draft for publication
- Public consultation – To be launched Nov/Dec 2025 for 3 months
- PRAC consultation
- Approval by BDSG
- Publication

Challenges - Dimensions

Challenges in ADR reporting organised by DQ dimensions

Reliability	Extensiveness	Coherence	Timeliness	Relevance
Accuracy Noise Incorrect assessment of causality	**Coverage** Underreporting	**Uniqueness** Duplicates	Delay with large (complex) veterinary cases	Cases from social media or e-commerce
Precision	**Completeness** Missing information Insufficient follow-up	**Format coherence** Different ways of reporting data Variability in data collection forms		
Traceability		**Semantic coherence** Different interpretation of regulations & grey areas Inconsistent coding		

Conclusion

Conclusion & next steps

- DQF draft has a number of relevant high-level recommendations
- Regulatory stakeholders have proposed detailed sub-recommendations to improve ADR reporting
    - Other stakeholders will be invited to propose further high-level recommendations and detailed sub-recommendations during the public consultation

Next steps

- Public consultation
- PRAC consultation
- Approval by NDSG
- Publication

EUROPEAN MEDICINES AGENCY
SCIENCE MEDICINES HEALTH

Thank you

tom.paternoster-howe@ema.europa.eu

Follow us

Recommendations

From the stakeholder workshop, interviews and regulatory consultation

1

Reliability

To what degree are data accurate or correctly representing an observed reality?

Accuracy (noise)

The amount of discrepancy between data and reality

- Recommendations:
  - Training: Enhance training for accurate causality assessment
  - Standardisation: Follow reporting guidelines
  - Data Entry: Improve structure in ADR data entry forms
    - Particularly relevant for spontaneous reporting: include drop-downs and auto-fill for text boxes and context-specific prompts and mandatory sections
  - Quality Control: Implement regular audits and feedback mechanisms
    - Feedback from users providing the data, those coding the received data and those reviewing the coded data

Precision

The degree of approximation by which data represents reality

- Recommendations:
  - Standardisation: use standardised reporting formats to enhance specificity
    - Include drop-downs and auto-fill for text boxes for data that is often irregular or misreported, e,g, batch numbers
  - Training: Conduct training to improve detailed/correct categorisation of ADRs
  - Quality Control: Implement regular audits and feedback mechanisms

Other DQ concepts related to Reliability

Plausibility: the likelihood of some information being true

Traceability: the knowledge of data source, and processing

- Recommendations:
  - Protocols: Follow reporting protocols to maintain the integrity of the original narrative
  - Version Control: Implement systems to track changes and maintain a history of modifications
  - Training: Educate stakeholders on the importance of traceability

2

Extensiveness

How much data do we have?

Completeness (Missing information)

Amount of information available with respect to the total available information

- Recommendations:
  - Standardisation: clear reporting requirements for ADR submissions
    - Ensure appropriate use (and not over-use) of pseudonymisation
  - Training: Enhance training for HCPs on the importance of completeness. Encourage detailed narratives in ADR reports written or finalised by humans and not only automatically generated
  - Follow-Up: Develop structured protocols for obtaining follow-up information
  - Technology: Utilise electronic systems to capture and verify data completeness
    - Include context-specific prompts and mandatory sections to capture key information in particular contexts (e.g. in pregnancy reports)

Coverage (Underreporting)

Amount of information available with respect to what exists in the real world

- Recommendations:
  - Awareness: Increase education on the importance of ADR reporting
  - Training: Enhance training on local protocols and reporting requirements
  - Culture: Foster a supportive environment for reporting ADRs
  - Engagement: Collaborate with stakeholders to promote reporting

Other DQ concepts related to Extensiveness

Representativeness: having the same characteristics as the whole it is meant to represent

Missingness: what is the impact of incomplete data in respect to coverage of a dataset

- Recommendations:
  - Define the target population and use stratified sampling
  - Assess patterns of missing data and implement new techniques

3

Coherence

Expresses how different parts of an overall dataset are consistent in their representation and meaning.

Format Coherence (Different ways of reporting data)

Whether data are expressed in the same way throughout a dataset

- Recommendations:
  - Standardisation: Use standardised reporting formats based on guidelines
  - Training: Enhance training on reporting standards for all stakeholders
  - Quality Control: Implement regular audits and feedback mechanisms

Structural or Relational Coherence

Whether the same entities are identified in the same way throughout a dataset

- Recommendations:
  - Audits: Conduct regular audits to identify and correct inconsistencies
  - Training: Enhance training for data handlers on coding
  - Quality Control: Implement regular audits and feedback mechanisms

Semantic Coherence (Different interpretation of regulations & grey areas)

Whether the same value mean the same thing throughout a dataset

- Recommendations:
  - Guidelines: Follow clear regulatory guidelines for ADR reporting
  - Standardisation: Establish a framework for consistent interpretation of guidelines
  - Collaboration: Foster cross-regional collaboration among regulatory bodies
  - Training: Enhance training for reporters on jurisdictional nuances
  - Quality Control: Regularly evaluate reporting practices for consistency

Uniqueness (Duplicates)

The same information should not be duplicated, but appearing in the dataset once

- Recommendations:
  - Standardisation:
    - Ensure clear rules are in place and followed to prevent replication
    - Follow standardised reporting protocols to minimise errors
  - Training: Enhance training for data entry personnel on possible faults
  - Review Process: Establish a systematic review process for detecting duplicates
  - Automated Sharing: Monitor automated data sharing to prevent replication

4

Timeliness

Are the data reflecting the intended reality at the point of time of its use?

Currency

How fresh the data are

- Recommendations:
  - Reporting Timelines: Respect deadlines for ADR reporting
  - Communication: Enhance communication channels between stakeholders
  - Follow-Up: Implement structured follow-up protocols for missing information
  - Training: Provide training on the importance of timeliness

5

Relevance

The extent to which a dataset presents the data elements useful to answer a given research question

Relevance

- Available data elements: Ensure that the standards for ADR reporting and collection continue to be pertinent for answering pharmacovigilance questions
    - Update international standards and databases as necessary to ensure that all the necessary data is available in a suitable format

6

Special considerations for specific ADR data sources

Patient support programmes, social media, e-commerce and OTC products

Patient support programmes, social media, e-commerce and OTC products

- Recommendations to be developed following publication of E2D(R1)
  - E2D(R1) recently published
  - Preliminary recommendations are being created internally and included in the version for publication

Recommendations

From internal stakeholder consultation

General recommendations from consultation with EMA signal assessors

These recommendations have been included in the previous slides as specific sub-recommendations within the overall framework

- Pregnancy – prompts/standardised FU for pregnancy reports, especially at the time of data entry from primary sources
- Narratives – ensure that autonarratives are reviewed by human case-processing staff, who are adequately trained & empowered to amended and rewrite the narratives to ensure coherence and readability
- Duplication – Stakeholders should cooperate to minimise the replication of ADR data
- Batch numbers – MAHs and Sponsors should ensure that their PV systems output accurate and consistent reporting of batch numbers for their own products – for example using standardised lists or format checks as part of the business rules for export

Dimensions discussed

Reliability	Extensiveness	Coherence	Timeliness	Relevance
Accuracy Noise Incorrect assessment of causality	**Coverage** Underreporting	**Uniqueness** Duplicates	Delay with large (complex) veterinary cases	Cases from social media or e-commerce
Precision	**Completeness** Missing information Insufficient follow-up	**Format coherence** Different ways of reporting data Variability in data collection forms		
Traceability		**Semantic coherence** Different interpretation of regulations & grey areas Inconsistent coding		