EUROPEAN MEDICINES AGENCY
SCIENCE MEDICINES HEALTH

Meta-data: data discoverability and data quality

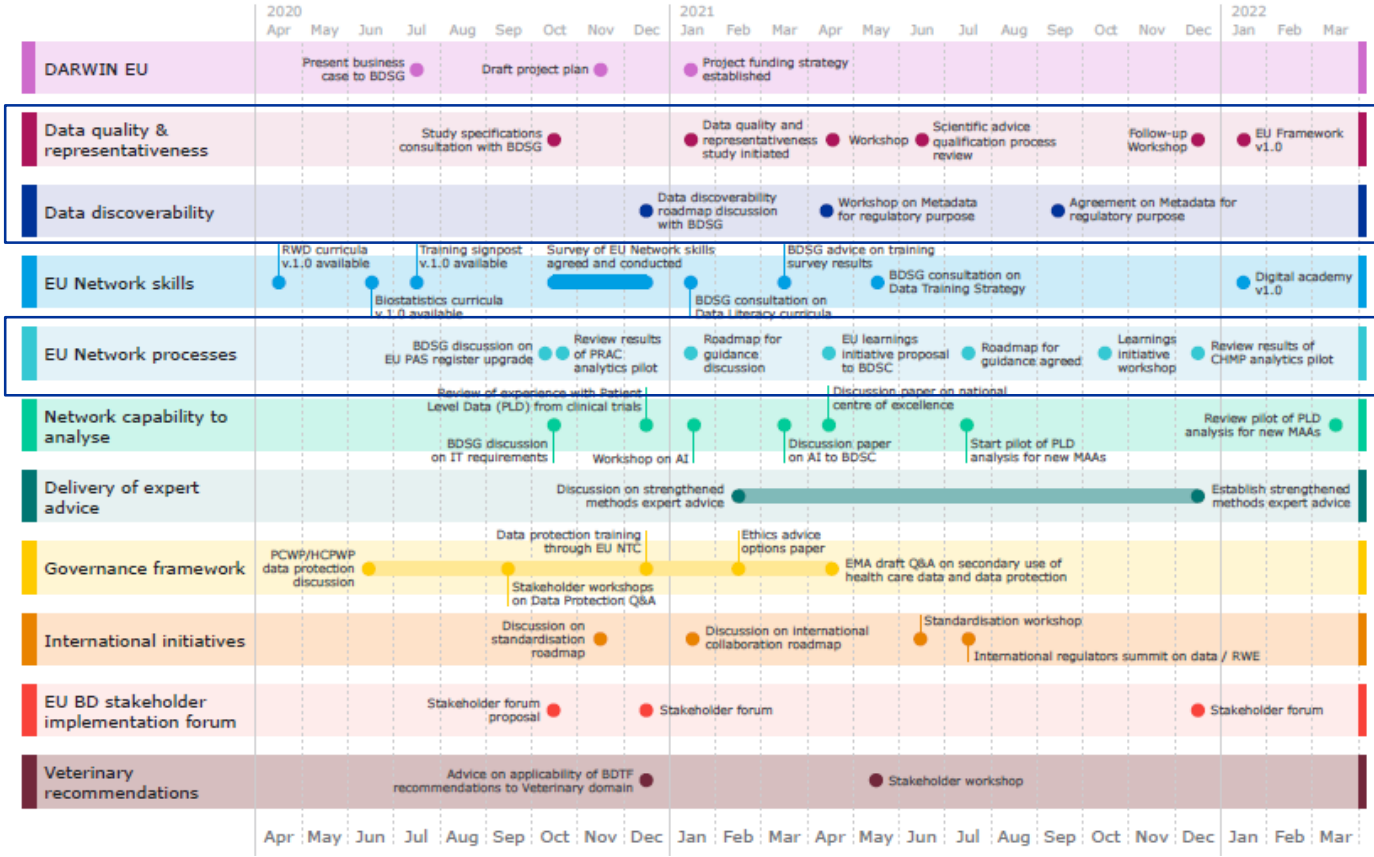
Agenda item 13

Presented by Katerina-Christina Deli (TDA-HCD)

Outline

- Big Data taskforce recommendations
  - Why they are important and potential applications
- Metadata, Data Quality framework and Catalogues Project
  - Data discoverability & EU resource database
  - Data quality
  - Transparency on study methods - EU PAS register

THE HMA-EMA JOINT BIG DATA STEERING GROUP WORKPLAN



HMA-EMA Big Data taskforce recommendations

Three recommendations strengthen the need to have a comprehensive knowledge of what data sources are available and their characteristics, and transparency on study methods

Recommendation 2: Establish an EU framework for data quality and representativeness

- Develop guidelines on data quality
- Strengthen the process for data qualification through Scientific Advice

Recommendation 3: Enable data discoverability

- Identify key meta-data for regulatory decision-making on the choice of data source
- Strengthen the current ENCePP resources database

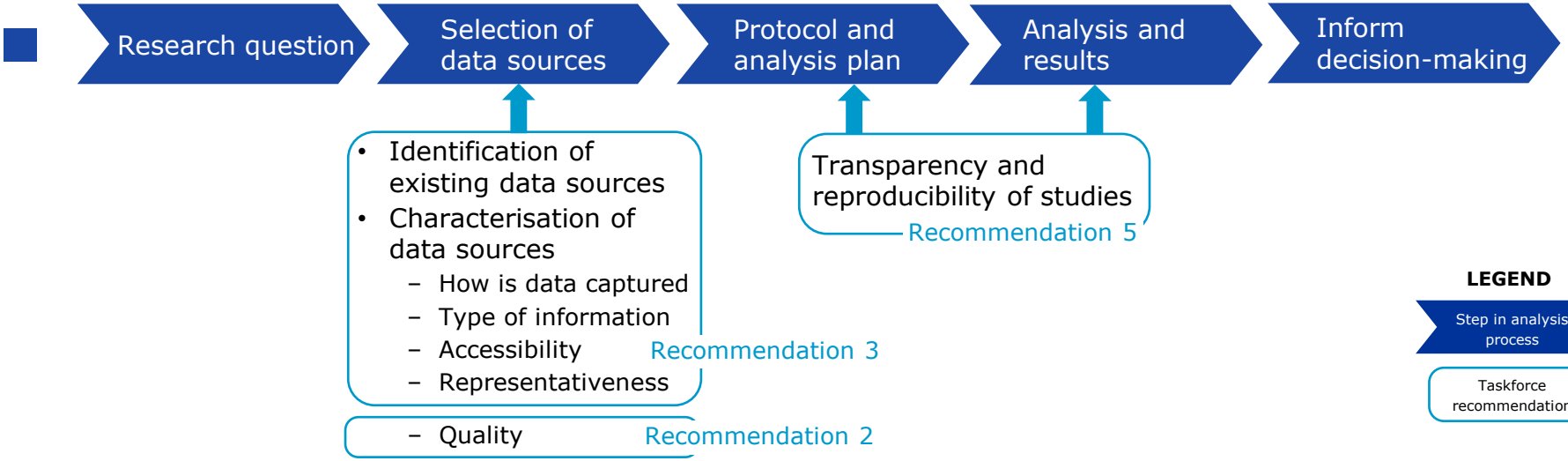
Recommendation 5: Strengthen EU Network processes for Big Data submissions

- Develop guidelines on study conduct and reporting
- Enhancement of the existing EU PAS register

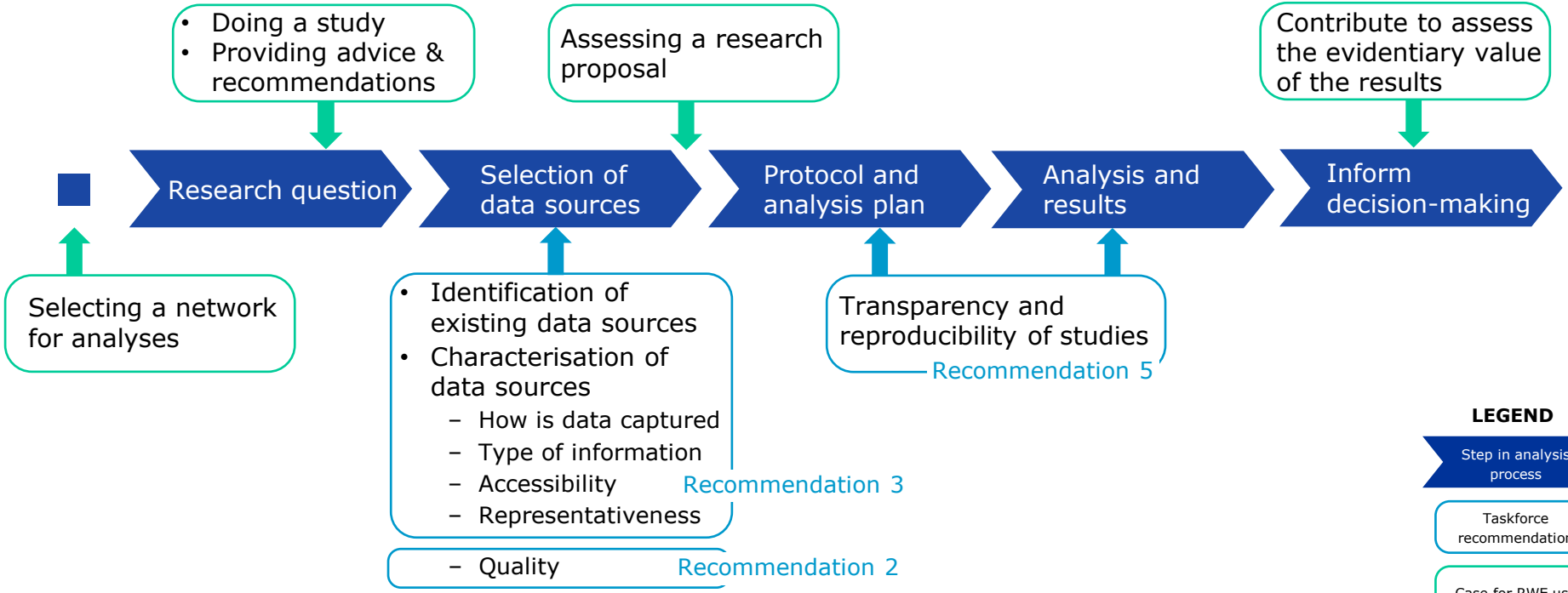
Current challenges

- Identification of appropriate real-world data sources is becoming an increasing need in regulatory decision making
  - Examples: long-term follow up of innovative medicines, post authorisation obligations for products authorised with a conditional authorisation
- Data needs are becoming more complex
  - Need for data sources of sufficient depth and details in several European Member States
- Lack of standardised information and statistics on real-world data sources
  - Data sets can be siloed by country, language, region, hospital and even department
  - Resource intensive to find suitable data sources, assess their characteristics and quality
  - Pharmaceutical companies may establish new data sources; duplication of effort and further fragmentation of the data landscape

Generating and assessing Real World Evidence to inform regulatory decision-making

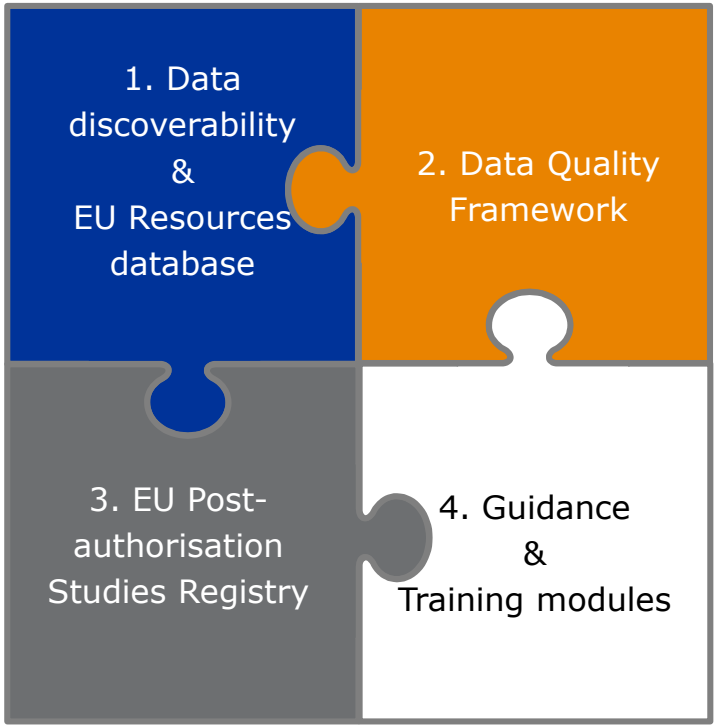


Generating and assessing Real World Evidence to inform regulatory decision-making



Metadata, Data Quality framework and Catalogues Project

- Identification of data sources
- Definition & collection of metadata
- Access (search, visualisation)



- Review of existing Data Quality Frameworks
- Establishment of a Data Quality Framework
- Understand data quality for regulatory purposes

- Collection of information on study protocols & results
- Collection of study metadata
- Access (search, visualisation)

- Identification of RWD sources for a specific study purpose
- Assessment of suitability of data sources used in studies
- Contribution in the assessment of the evidentiary value of study results

Data discoverability – Data sources

Real-world databases to be included are mainly:

- Databases allowing to link drug utilisation data to subsequent and existing clinical events and demographic variables for individual patients:
  - Primary care, specialist care, hospital care data from EHRs
  - Claims databases
  - Disease registries
- Databases allowing to measure for each individual patient the duration of use and the cumulative doses of medicines prescribed/delivered:
  - Longitudinal drug prescription
  - Dispensing
  - Other drug utilisation

Data discoverability - Metadata

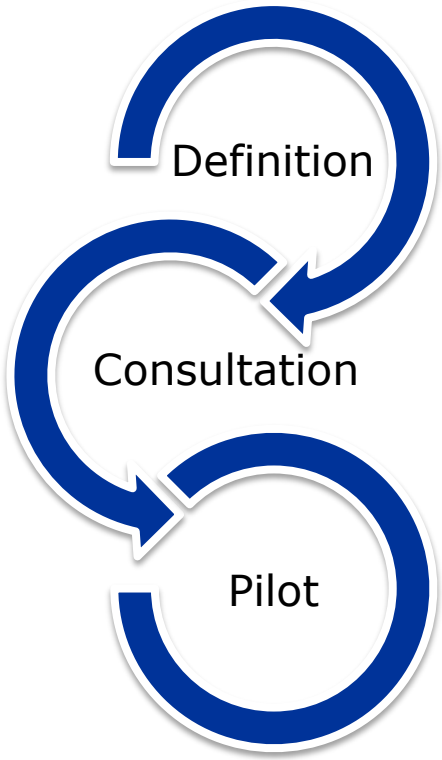
- "Metadata are descriptive data that characterise other data to create a clearer understanding of their meaning and to achieve greater reliability and quality of information"*
- Categories of metadata (examples):
  - General (size, type of data, follow up, distribution of the population by key characteristics, linkage, dictionary, terminologies used for coding, use of common data model)
  - Technical (lag time)
  - Governance (accessibility, ethics approval)
  - Quality (including publications)
  - Representativeness of the population covered (e.g. age, gender, geographical distribution, frequency of diagnoses)

*American Health Information Management Association (AHIMA), *Pocket Glossary of Health Information Management and Technology*. 2012, Chicago, IL: AHIMA Press.

Data discoverability – EU Resources database

- What is it
  - Establishment of a database and related dashboard of real-world data sources
  - Upgrade of the current ENCePP Resources Database or development of a new database
  - Publicly available with search and visualisation functionalities of key metadata related to the content, quality and representativeness of the individual real-world data sources
- Why is it important
  - Help to leverage the richness of the available EU data landscape to inform regulatory decision-making
  - Promote the registration and identification of datasets, centres and networks
- How will it be maintained
  - Sustainability plan will be developed including clear and concrete incentives for data holders to ensure the maintenance of high-quality information

Data discoverability – steps and timelines

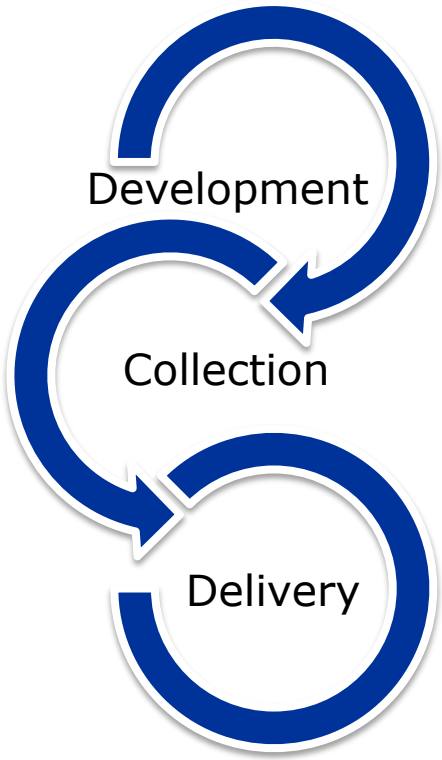


Definition of set of metadata relevant for regulatory decision making from real-world data sources looking at metadata used by other regulators or organisations

Stakeholder engagement, consultation and workshop are part of the work plan for 2021

Delivery of the final list of metadata by the end of 2021. The metadata collection will be piloted for a limited set of data sources (covering different type of RWD sources and formats). Delivery of a good practice guide on data discoverability (including description of the metadata and advice on use) by the end of 2021

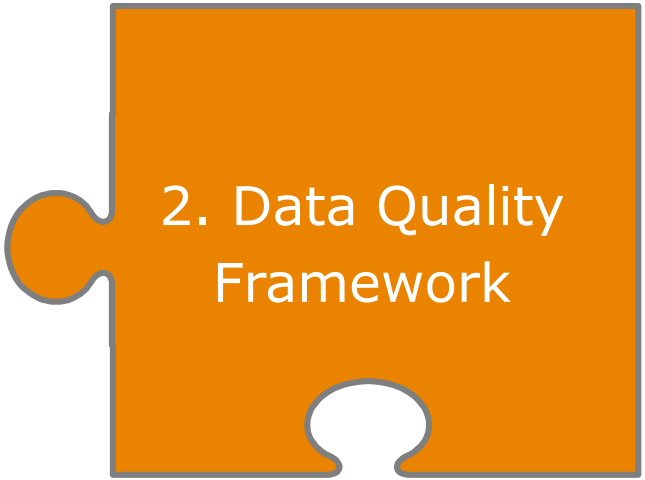
Data discoverability– steps and timelines



Development of the EU Resource database and related dashboards (date to be confirmed)

Identification of all the databases to be included in the EU Resource Database catalogue and collection of metadata from all the databases (date to be confirmed)

Delivery of completed EU Resource database (date to be confirmed)



Development of a Data Quality Framework

Define quality standards for source data and transformed data. Range of applicability, quality control measures and limitations of data

With the measurements in place, a classification system needs to be created, defining minimal requirements tailored for each intended regulatory purpose

Quality standards – guiding concepts (examples)

Conformance

Values are in the intended format and allowed values

Uniqueness

Measures of unwanted duplication, within or across databases

Completeness

Frequencies of missing values

Accuracy

Correctness of the content of the data

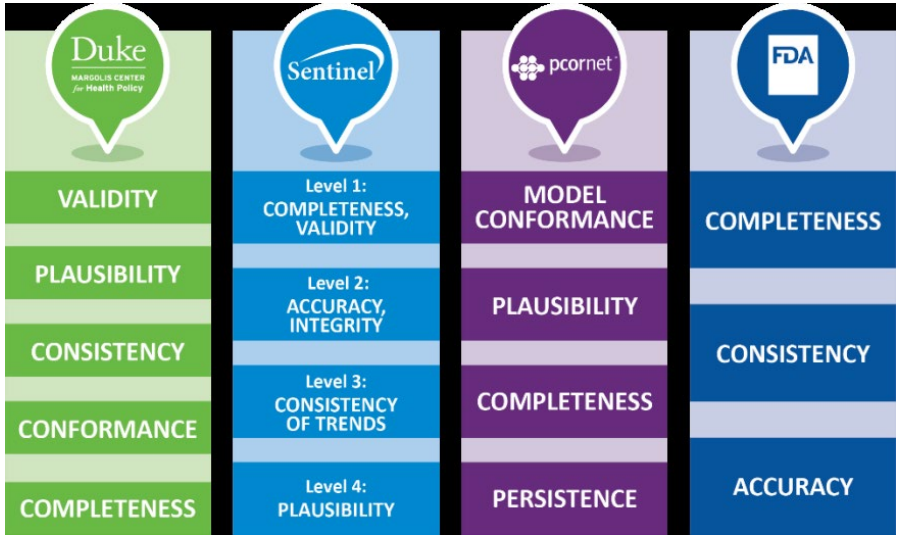
Atemporal plausibility

Data values or distributions agree with common knowledge or with gold standards

Temporal plausibility

Time varying variables change values as expected

Revision of existing data quality frameworks



The revision aims to establish harmonized data quality concepts, methods, and to establish a common understanding of the strengths and limitations of data for regulatory purposes

Source: Determining Real-World Data's Fitness for Use and the Role of Reliability

Drafting a Data Quality framework – steps and timelines

Comprehensive revision of existing data quality framework initiatives will be carried out looking at previous experience such as EHDEN, FDA Sentinel, Health Canada CNODES, and the Observational Health Data Sciences and Informatics (OHDSI) open-source DQ tools

Stakeholder engagement and consultations and a workshop are part of the work plan for 2021

Delivery of first draft of the Data Quality Framework. Development of guidelines and training modules related to practical aspects of data quality measurements including expert's recommendation based on the collected metadata, data quality measurements, use of data and its limitations (date to be confirmed)

Transparency on study methods – EU Post-Authorisation Studies Register

- What is it
  - Establishment of a global repository and related dashboard with information on studies protocol and results
  - Will include metadata on studies to support identification and comparability of studies
  - Upgrade of the current EU PAS Register or development of a new database
  - Publicly available with search and visualisation functionalities
  - Relevant links to be established with other tools (EU Resources Database, CTIS, SPOR)
- Why is it important
  - To promote exchange of information on observational research among stakeholders (including academia, sponsors and regulatory bodies)
- How will it be maintained
  - Automatic checks/follow-up alerts will support up to date information
  - Sustainability plan will be developed including clear and concrete incentives for investigators to ensure the maintenance of high-quality information

Any questions?

Further information

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

Follow us on **@EMA_News**