Stakeholders feedback from pre-meeting survey: Session 2 Metadata

Technical workshop on real-world metadata for regulatory purposes
Virtual meeting, April 12, 2021

Presented by Dr. Romin Pajouhesnia
Utrecht University

Survey questions

1. Are there metadata variables required for characterising real-world data sources and the regulatory use cases they will support that you would like to see included in the catalogue, which are not included in the current preliminary list?

   If so, can you provide your suggestion(s) with a description of the meaning(s) of the variable(s)?

2. Please share any questions or comments relating to the preliminary definitions, structure of the proof-of-concept catalogue and its organisation into six key domains: *institution, data source, data bank, common data model, network, study*.

Feedback

- Q1: On required RWE metadata variables not included
  - OK/none/blank/provided comments in interview: 10
  - We appreciate the amount of work that has been done by the team.

- Q2: Preliminary definitions, structure, organisation into six domains
  - OK/blank 8

All feedback will be taken into consideration when creating the final metadata list

General comments

- **Purpose of the catalogue**, "…currently mixture of a database catalogue, competence catalogue, network catalogue, study catalogue…etc.", "…focuses on Clinical Trial more than on regulatory activities…"

Clarification: Aim to describe data sources in sufficient detail to support use of RWD in regulatory activities. Focus is on sources of routinely collected data (RWD) – clinical trials not within scope.

- **Definition of fit-for-purpose**, "…the idea of appropriateness requires thorough fit-for-purpose assessments, which could require data, technology and knowledge."

Clarification: Very important. Assessment of suitability of the metadata list for specific use cases is planned

- **International standards** need to be incorporated –CDISC, HL7 FHIR standards, GVP Module V, SPOR.

Actions: Fully intend for the final metadata list to be adherent to existing standards – this will be described in more detail in sessions 3 and 4 this afternoon. Where needed, concepts may be submitted as new standards.

- **Richer code vocabularies** – ICD versions and modifications, MedDRA, LOINC, ISBT-128

Actions: Vocabularies will be expanded; could be introduced during process of populating the catalogue

Additions and clarifications: Institution, Network

Comments on the terms used and the roles of institutions

- "…ensure a better understanding of who is responsible for what.","…seems to include all institutions listed in the study table. Is it for referencing the research centers familiar in analyzing/using such data?"

Clarification: Intended that metadata on institutions that contribute to the catalogue is collected:

- who is responsible for catalogue entries, expertise/experience, relation to data sources (e.g. access, maintainer)

- Category "institutions" as "data users" or "data contributors"; clarification of *data access provider* compared to *data user* and *data custodian*

Clarification: We acknowledge the need for clearer description of the different roles of different institutions

- critical for understanding relation to data sources

Additions and clarifications: Data source and Data bank

- Need for richer metadata variables and definitions describing access and permissions

Action: Additional metadata variables need to be added. Challenging to distil such a set.

- Include data bank within data sources?

Clarification: This separation was proposed for reasons explained in Session 1.

- Additional metadata variables suggested
- The revised list will clarify that quantitative measures over time (population size etc.) are included in the list
- Valuable suggestions for additional metadata on products, medical, sociodemographic information, fields capturing qualification of data sources
- *Data bank - Prompts* needs further explanation – it addresses a number of the received comments
- Quantitative descriptors of completeness and counts e.g. disease cases (where available) can be enriched
- *Study – data characterization* table captures quantitative descriptors including conformance and plausibility

Additions and clarifications: Study and CDM

Comments on Studies tables

- "…includes in itself a universe" – suggestion for inclusion of unified taxonomies
- Relation to EU PAS register?
- Suggested metadata: standard descriptions, study rationale, differences between data sources in a study
- "proper standardization" of uploaded data characterization documents needed.

Clarification: We agree. This will be investigated in the next steps of the project

Comments on Common data models tables

- "Leave CDMs to be defined elsewhere", "better handled by the respective community",

Clarification: Intended to describe existing mappings. CDM details is indeed best captured by links

- "How would this be useful for a user of the catalogue?"

Clarification: Envisioned to provide a information on availability of CDMs as a tool to support secondary use of RWD. Existing mappings can be useful in study planning (e.g. feasibility of rapid assessment).

Structure of the preliminary metadata list

Comments on how the different domains are envisioned within the same catalogue

- "…different requirements in terms of accuracy, completeness, timeliness, uniqueness etc. … have their own dynamic and are complex digital objects of their own."

Clarification: Pertinent to the design of a process to gather and maintain metadata – the catalogue could be flexible in this. More to be explained in session 3 this afternoon.

Comments on table structure

- "Overall, the 6 domains make sense, however, more structure would be beneficial for the detailed information included in DS/DB, e.g. in form of sub-domains…"

Action: Helpful suggestions provided for more user-friendly organization of metadata variables in a number of tables will be strongly considered when finalizing the list

Are we asking for too little or too much?

- “Some of the attributes are more than meta-data in our perspective, so perhaps the catalogue is **too "big".**”
- “Think what can be sacrificed” “Is there a **minimal set of fields** that would be prioritized for a particular database record to be considered 'complete’?”

Take home: Balance between completeness and practicality

Actions: Continue to propose a set of priority (or mandatory) fields. Tools required to support extraction of metadata

Thank you!

For any question on this presentation, please contact: Malgorzata.Durka-Grabowska@ema.europa.eu
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