EU Real World metadata list, good practice guide and catalogues of data sources and studies

EMA HMA Big Data Stakeholder Virtual Forum – 1st December 2022

Presented by Katerina-Christina Deli

Healthcare Data Workstream – Data Analytics and Methods Task Force

Outline

1. Priority recommendation – data discoverability
2. Achievements and outlook for 2023 and beyond
3. Metadata list & Good Practice Guide – overview & comments from public consultation
4. Status update on development of the catalogues
5. EU Metadata catalogue mock-ups

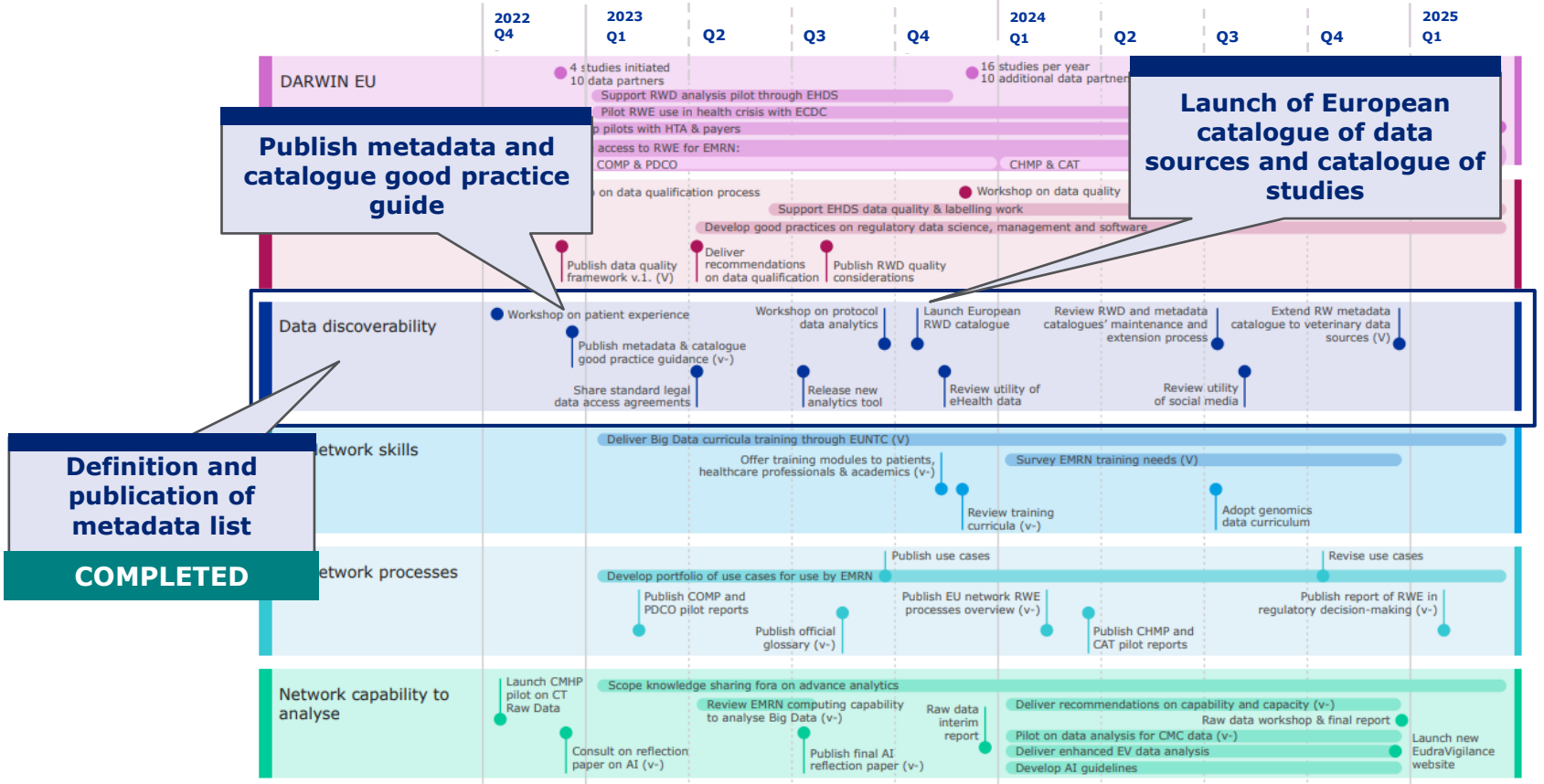
Reminder – data discoverability

Goal of setting up an EU metadata catalogue of data sources and a catalogue of studies

- Facilitate the **discoverability of data sources** to generate adequate evidence for regulatory purpose
  - initial identification of data sources suitable to investigate a specific research question, which could be further explored as to the accessibility and availability of the required data

- Support the **assessment of study protocols and study results** by providing quick access to information on the suitability of data source(s) proposed to be used in the study protocol or referred to in the study report

BDSG Workplan – data discoverability milestones



Achievements and outlook for 2023 and beyond

2021

MINERVA pilot study

Conducted by a consortium including 16 organisations from 9 countries

Definition of first draft of metadata list and engagement with stakeholders

Metadata and practices of multiple key regulators and organisations were considered as potential sources of information

Stakeholder workshop April 2021

Options for the process to collect and update the metadata

Catalogue proof of concept tool for accessing, searching, and visualising the metadata collected

2022

Surveys consulting on content and functionalities of the catalogues within ENCePP community, industry, international regulators

Webinar involving TEHDAS, NCA representatives (via EUNDB, BDSG) to consult on RW metadata list

Agreement on RW metadata list for regulatory purposes

Publication of metadata list

Development of EU metadata catalogue started

Data collection from different data sources started

Public consultation of good practice guide on metadata collection

2023 onwards

Publication of good practice guide on metadata collection

Migration of data from ENCePP and data enrichment

Data collection from different data sources (DARWIN EU®, collaboration with other catalogue initiatives) to be continued

Launch new EU metadata catalogue

Maintenance of RW metadata catalogue

Real-world metadata list and good practice guide

Metadata list for real-world data

- Describes **real-world data sources and studies** to help pharmaceutical companies and researchers to **identify** and **use** such data when investigating the use, safety and effectiveness of medicines
- Will **feed** into two future EU catalogues:
  - **Catalogue on real-word data sources** which is currently being built and will replace European Network of Centres for Pharmacoepidemiology and Pharmacovigilance (ENCePP) catalogue in late 2023
  - **Catalogue on real-world data studies** which will cover studies performed on the data sources, enhancing and replacing the European Union electronic register of post-authorisation studies (EU PAS Register)

Good Practice Guide for the use of real-world metadata

In anticipation of the real-world data sources catalogue launch in 2023

- Provides **recommendations** to a variety of stakeholders for the **use** of the catalogue to **identify real-world data sources** for **assessing** the **suitability** of data sources for specific studies
- Describes the **metadata elements** that are envisaged to be used in the future EU catalogue of real-world data sources

Good Practice Guide – Scope and Table of Contents

Scope

- **Use of the catalogue** to identify real-world data sources to assess the suitability of data sources for specific studies

- **User Guides**
  - **Metadata definitions** and information expected to be used when describing a data source
  - Guides the user on **adding new data and maintenance** of data in the catalogue (to be further expanded once the system is live (e.g.: screenshots, technical guidance)

Contents

Assessing suitability of data sources used in studies

- **Two important aspects** of data quality of the primary data:

Reliability

Based on e.g. the detection and correction of errors, missing data and implausible values (characteristic of the data source independent from its use for a specific study)

Relevance

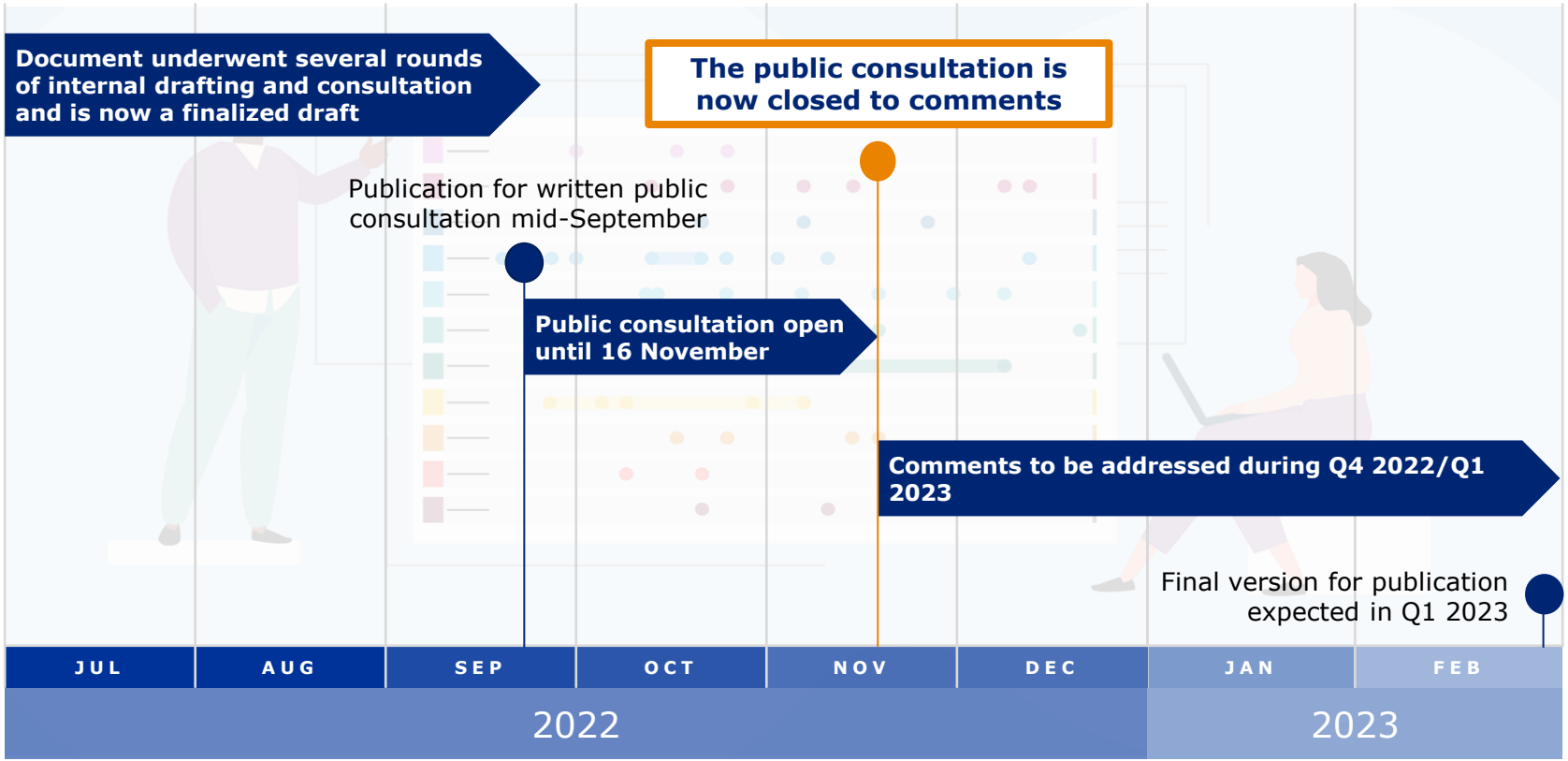
Data source to provide adequate and valid evidence informing a specific research question following the application of appropriate epidemiological and statistical techniques (may be required for some studies and not for others)

- The good practice guide describes the **specific variables collected** (metadata) to be used to specifically evaluate the data source's reliability and relevance

Use cases described in the draft Good Practice Guide

Use case	Content
An investigator wants to identify suitable data sources for a planned study	Outline of the steps recommended for a catalogue user in order to select a suitable data source for a particular study
A study protocol submitted for a study mentions a data source and the assessor needs to understand the suitability of the data source proposed	Description of the data elements useful for the assessment of the suitability of data source in the assessment of a study report
An investigator writes a study reports for which they need to describe the data source(s) proposed to be used or used in the study	Proposal of solutions when a comparison between the characteristics of several databases used in the study is difficult to perform
A data holder wants to compare the features of a specific data source with other ones covering fully or partially the same population	Recommendations on benchmarking several data sources – use of the harmonised description of the characteristics of each data source

Next steps – GPG drafting and publication



Some comments from the public consultation

We received contributions from regulatory network, data holders, researchers and research networks, industry, medical associations, etc.

Thanks to all who contributed!

Comments were mainly related to:

Catalogue maintenance and content validation

- Incentivisation of data holders to provide and update the metadata
- Quality check of data collected (how/when) to ensure data are entered in a harmonised way
- Management of the evolvement of metadata elements collected over time

Scope of data source catalogue

- Clarify whether the catalogue will accommodate data sources outside the EU/EEA
- Clarify scope in terms of type of data sources (e.g., registries, electronic health records, claims data, m-health) and if there are exceptions

Interoperability/collaboration with other catalogues (e.g., DARWIN EU®, EDHEN)

- Harmonisation of definitions and metadata terminology
- Minimisation of effort duplication for data holder

Importance of data source accessibility

- In the process of identifying suitable data sources for a study this should be an early step
- Add in the catalogue information on accessibility, timelines and user restriction

Interlink of Data quality framework (DQF) with the data sources catalogue and Good practice guide

- Alignment between DQF and Good practice guide terminology and definitions
- Application of the DQF to the catalogue

Need for clarifications/definitions for specific metadata or proposals more metadata to collect

- Include more standardised vocabularies or use additional terminologies (e.g. to describe disease information)
- Definition needed (e.g. for active population)
- Collection of more quantitative metadata / Producing quantitative descriptive statistics

Status update on the development of the catalogues

Discovery and analysis phase of the project to conclude by end of 2022

Testing to include **input from key stakeholders** (e.g., industry, ENCePP)

Launch is planned for Q4 2023 / Q1 2024

Key features:

The **study catalogue and the data sources catalogue** will be linked → possibility to view information about data sources used in the study and vice versa	User **registration and login** will be required
Possibility to **search and export** information	External users will be able to **submit data through a webform** (insert/update content)
Studies will be visible in the relevant **medicines overview page** in the EMA website	The approval/rejection of data will trigger **e-mail notifications to user submitting data** (i.e.: the EMA administration functions)
Globally unique and persistent identifiers according to FAIR (Findable, Accessible, Interoperable, and Re-usable) data principles	**Notifications will be sent prompting users** to update their records

Thank you for listening

Further information: metadata@ema.europa.eu

See websites for contact details

Heads of Medicines Agencies www.hma.eu
European Medicines Agency www.ema.europa.eu

The European Medicines Agency is an agency of the European Union