Metadata, Data Quality framework and Catalogues Project

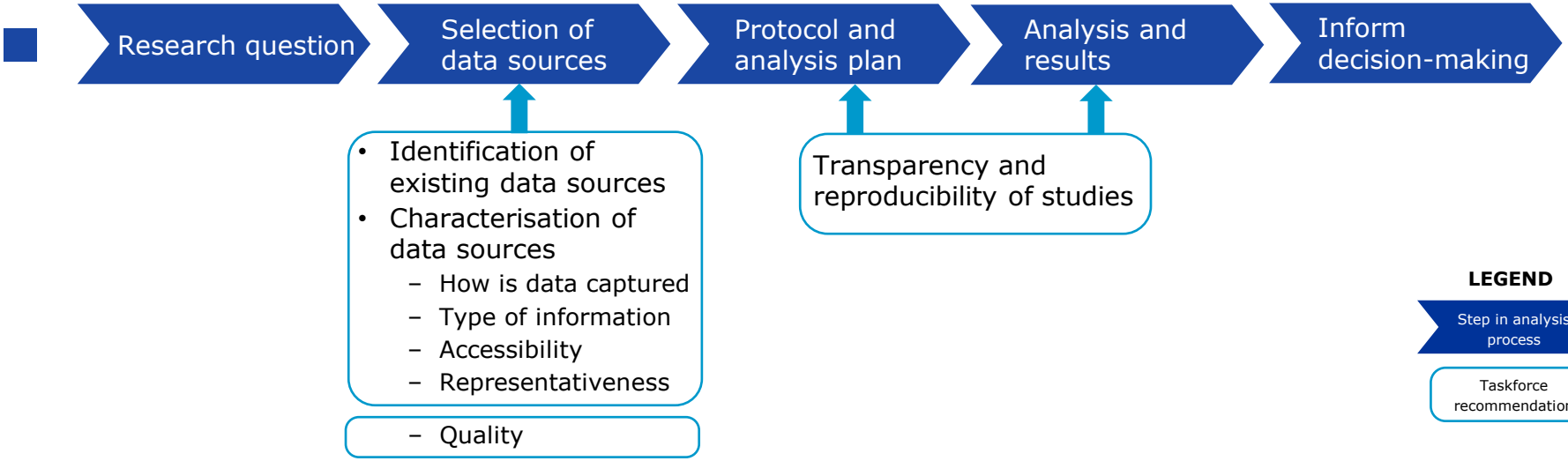
Presented by Xavier Kurz– 18 November-2021 ENCePP Plenary meeting

Background and scope

EUROPEAN MEDICINES AGENCY

Report of the HMA-EMA Big Data Task Force Phase II – January 2020

Process for generating and assessing Real World Evidence to inform regulatory decision-making



- 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
- Lack of standardised information and statistics on real-world data sources
  - 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

This project aims to bring about **enhanced transparency** with regards to observational studies and data sources in the regulatory context and increased quality of these, **enhanced discoverability** of studies and data sources and the **ability to evaluate the level of evidence** provided by observational studies and real-world data sources when used in the regulatory context

Current status: ENCePP Resource Database

EUROPEAN MEDICINES AGENCY

Created 10 years ago by ENCePP, it provides an unprecedented level of knowledge and transparency on sources of data for pharmacoepidemiology in Europe. Some improvements are needed to match the expansion of the pharmacoepidemiology field in the last decade.

Data fields

The current data fields: these are insufficient to identify key characteristics (meta-data) to inform on the relevance of the registered data sources

Data maintenance

Maintenance of the registered data: From the 143 data sources registered as of May 2020, more than 70% (104) have not been updated in the last 2 years, including decommissioned sources

Functionalities

Search and export functionalities: currently, the possibility to search and export is minimal, limited only to a few structured fields, which reduces the discoverability of registered data sources

It is difficult to link data sources and studies performed using them

[3] Metadata = "data about data"

Current status: EU PAS Register

EUROPEAN MEDICINES AGENCY

Created by ENCePP primarily to exchange information between centres and provide transparency on studies, the ENCePP eRegister of studies was the first public register of observational studies. Adopted as the EU PAS Register foreseen in the 2012 pharmacovigilance legislation.

Data fields

The current data fields should be updated and tailored to the scope of the studies. At the moment, investigators are required to fill-in information which may not ease the identification of relevant study characteristics. Some fields of study categories are outdated.

Data maintainance

Maintenance of the data registered: 30% of the registered studies marked as "ongoing" or "planned" have not been updated according to the "planned finalisation dates" as entered in the database ranging from 2012 to 2019.

Functionalities

Search and export functionalities: the possibility to search and export is minimal and limited to a few structured fields only, which restricts the discoverability and comparability of information related to registered studies.

Rebuilding of catalogues

EUROPEAN MEDICINES AGENCY

User stories

- Who uses the catalogues, how and to what purpose

Starting point; some of the areas below will stem from here

Data fields

- Starting point: current metadata list proposed by MINERVA project
- Define mandatory vs optional elements

Data maintenance

- Mapping of current existing fields to the new fields
- Migration of current data to new format
- Way forward for missing information

Functionalities

- User access and data input
- Search functionalities
- Export functionalities
- Usability
- Interconnection between catalogues
- Dashboards, visualisation
- Data quality elements

Metadata, Data Quality framework and Catalogues Project

1-year-study (MINERVA):
- ✓ Criteria for database selection
- ✓ Common set of metadata for describing real-world data sources
- ✓ PoC catalogue with min 10 databases

2+1+1 year-study:
- ✓ Data Quality Framework
- ✓ Collection of metadata from eligible data sources

↔

Catalogues:
- ✓ Catalogue of Real-World data sources
- ✓ Catalogue of observational studies

↓

A Partnership of the EU PE&PV Research Network, the SIGMA Consortium, and Collaborators

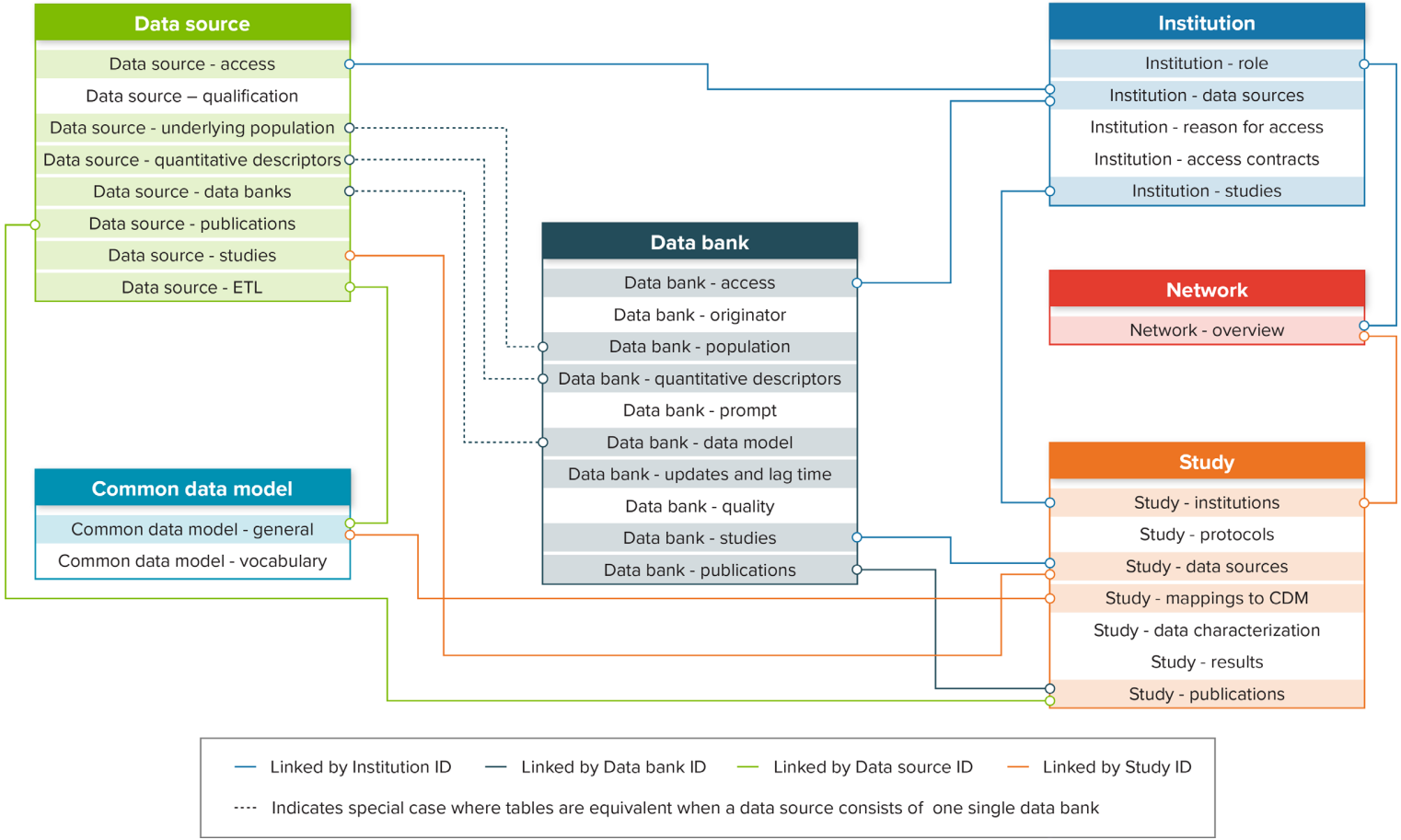
EUPAS 39322

1st Step: MINERVA

- One year study initiated in November 2020
- Contracted to RTI Health Solutions, which has formed a consortium including 16 organisations from 9 countries
- Main outcomes:
  - Definition of the set of metadata for identification of data sources and description of their characteristics in order to:
    - Select suitable data sources to address specific regulatory use cases
    - Assess data sources proposed in studies
    - Contribute to the assessment of the evidentiary value of study results
  - Pilot the process of collecting the data sources and their metadata including the creation of a tool to collect, search and visualise the metadata

Rebuilding of catalogues (MINERVA)

EUROPEAN MEDICINES AGENCY



2nd Step: 2+1+1 study: Data Quality Framework & Metadata

- Tender published in 2021 – contract finalisation
- Activities will be organised around two main work areas:

Work area 1: Data Quality Framework

- Revision of past experience and other existing data quality frameworks
- Drafting a Data Quality Framework with a starting point in the existing knowledge
- Workshop consultation with stakeholders
- Application of the framework in practical scenarios and use cases to test its usability

Work area 2: Metadata

- Enrich the data source catalogue with more data sources
- Good practice guide: focus on regulatory use cases and practical examples
- Identify metadata for observational studies catalogue

2+1+1 study: Data Quality Framework

The data quality framework should address:

- Data quality principles (e.g. drafting of best practice guides, procedures to follow, understanding data needs);
- Data quality dimensions (e.g. completeness, uniqueness, timeliness, validity etc.);
- High-level principles and definitions applying to all data types;
- Data quality standards related to metadata (e.g.: including measurements and address the integration with ISO standards referring to data quality, Information technology — Big data — Overview and vocabulary, CDISC, etc.);
- Communication guidelines on clarity and transparency principles for data quality issues;
- A series of applied use-cases and examples for regulatory purposes.

3rd Step: Rebuilding of catalogues

Catalogue of data sources

- Envisaged to contain information about available databases (discoverability)
- Information listed on each database based on the metadata defined in the MINERVA and 2+1+1 project
- A subset of the databases may have a more advanced visualisation/detailed information (e.g.: dashboards, graphs, side by side visualisation) -> "fingerprinted catalogue"
- Replacing the current ENCePP Resource Database catalogue

Catalogue of studies

- Linked with the data sources catalogue to provide information about studies performed in data sources and about data sources used in studies
- The metadata needs to be defined (based on MINERVA's proposal)
- Replacing the current EU PAS Register

Any questions?

Further information

[Insert relevant information sources or contact details as applicable.]

Official address Domenico Scarlattilaan 6 ● 1083 HS Amsterdam ● The Netherlands
Telephone +31 (0)88 781 6000
Send us a question Go to www.ema.europa.eu/contact

Follow us on **@EMA_News**