EUROPEAN MEDICINES AGENCY
SCIENCE MEDICINES HEALTH

HMA-EMA Catalogues of real-world data sources and studies

Facilitating Data Discoverability and Advancing Transparency of Real-World Evidence in EU Medicines Regulation

Multi-stakeholder workshop on Pharmacogenomics
24 September 2024

Presented by Stefania Simou
European Medicines Agency - Data Analytics and Methods Taskforce

Background & Context

The European Medicines Agency (EMA) and Heads of Medicines Agencies (HMA) set up a **joint task force to describe the big data landscape** from a regulatory perspective and **identify practical steps** for the European Medicines Regulatory Network to **make best use of big data in support of innovation and public health** in the European Union (EU).

The **HMA-EMA joint Big Data Task Force**, also known as the **Big Data Steering Group**, was established in December 2018. It developed **Ten Priority Recommendations** to advance the use of big data in the European regulatory network and advises EMA and HMA on prioritisation and planning of actions to implement them.

Recommendation 3:
Enable data discoverability

- Identify **Key Metadata for regulatory decision-making** on the choice of data source
- **Strengthen the ENCePP** Resources Database and **EU PAS register**

HMA-EMA Catalogues of real-world data sources & studies

EUROPEAN MEDICINES AGENCY

Public metadata repositories that describe **RWD sources and studies** that utilise such data to generate RWE. The Catalogues help regulators, pharmaceutical companies and researchers **identify and use** such data when investigating the use, safety and effectiveness of medicines.

Launched February 2024

Catalogue of RWD sources

replaces the European Network of Centres for Pharmacoepidemiology and Pharmacovigilance (ENCePP) Resources Database

Catalogue of RWD studies

replaces the European Union electronic register of post-authorisation studies (EU PAS Register®)

https://catalogues.ema.europa.eu/

- Enhancing **discoverability and evaluation of data sources and studies** -> facilitating the use of RWD sources, ultimately supporting evidence-based regulatory decision-making
- Facilitate **collaboration and research**
- **Link between RWD sources and studies** which can support **study design**, **protocol evaluation**, and **results interpretation**
- Promote **transparency** in observational research and **reduce publication bias** by making publicly available metadata on non-interventional studies
- Promotion of good practices aligning with **'FAIR' data principles** for **F**indable, **A**ccessible, **I**nteroperable, and **R**eusable data
- Respond to the **DARWIN EU open call** to become a DARWIN data partner via the Catalogues
- **User-friendly platform** for researchers, regulators, pharmaceutical companies and general public
- **Freely available access** via the Catalogues webpage, hosted on EMA public website

RWD Catalogues – Timeline

(**2020**) The Big Data Steering Group (BDSG) workplan **Recommendation 3: Data discoverability**

(**2021**) Data discoverability multistakeholder **workshop** followed-up by wide-reaching **surveys**

(**2022**) Publication of [Good Practice Guide on the use of the RWD Catalogue]

(**2024**) **Launch of HMA-EMA RWD Catalogue**

(**2021**) **MINERVA project** sets out the foundations of the activity in a one-year preparatory work

MINERVA: Metadata for data dIscoverability aNd study rEplicability in obseRVAtional studies

Strengthening Use of Real-World Data in Medicines Development: Metadata for Data Discoverability and Study Replicability (EMA/2017/09/PE/16)

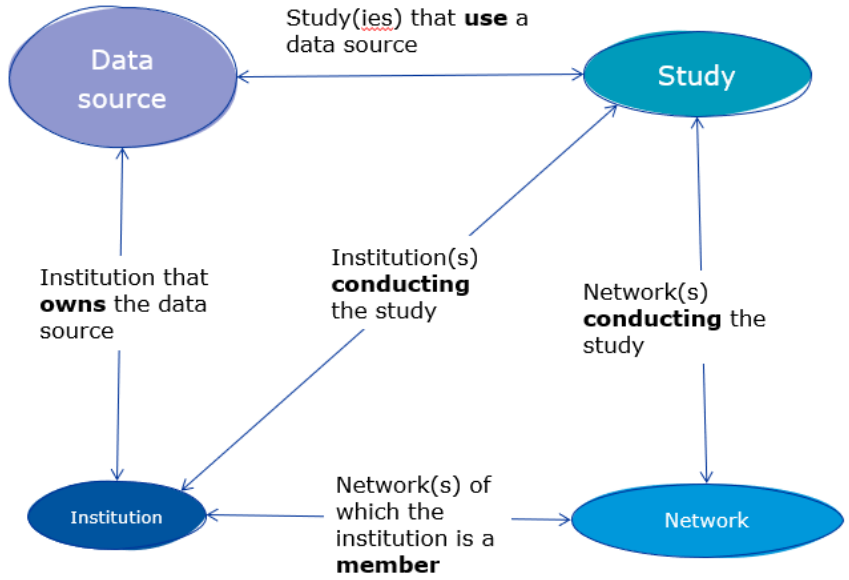
(**2022**) Adoption and publication of the "[List of metadata for data discoverability]"

(**2023**) **Migration** of ENCePP data sources and studies. Metadata collection for data sources of interest. Development of a new IT system.

Data collected in the Catalogues

The EMA-HMA have published the List of metadata for Real World Data catalogues, in which the main types of defined Metadata are:

Data Sources	Metadata containing the details of the data source (e.g., data source countries, data elements collected, etc.)
Studies	Metadata describing studies conducted using data sources described in the catalogue, EU PAS Register
Institutions	Metadata on any contributor to the catalogue, its role and expertise (e.g., institution country, etc.)
Networks	Metadata describing networks/consortia linking to institutions, data sources and studies in the catalogue (e.g., network name, website, etc.)



Heat map of data sources & studies registered in RWD Catalogues

EUROPEAN MEDICINES AGENCY

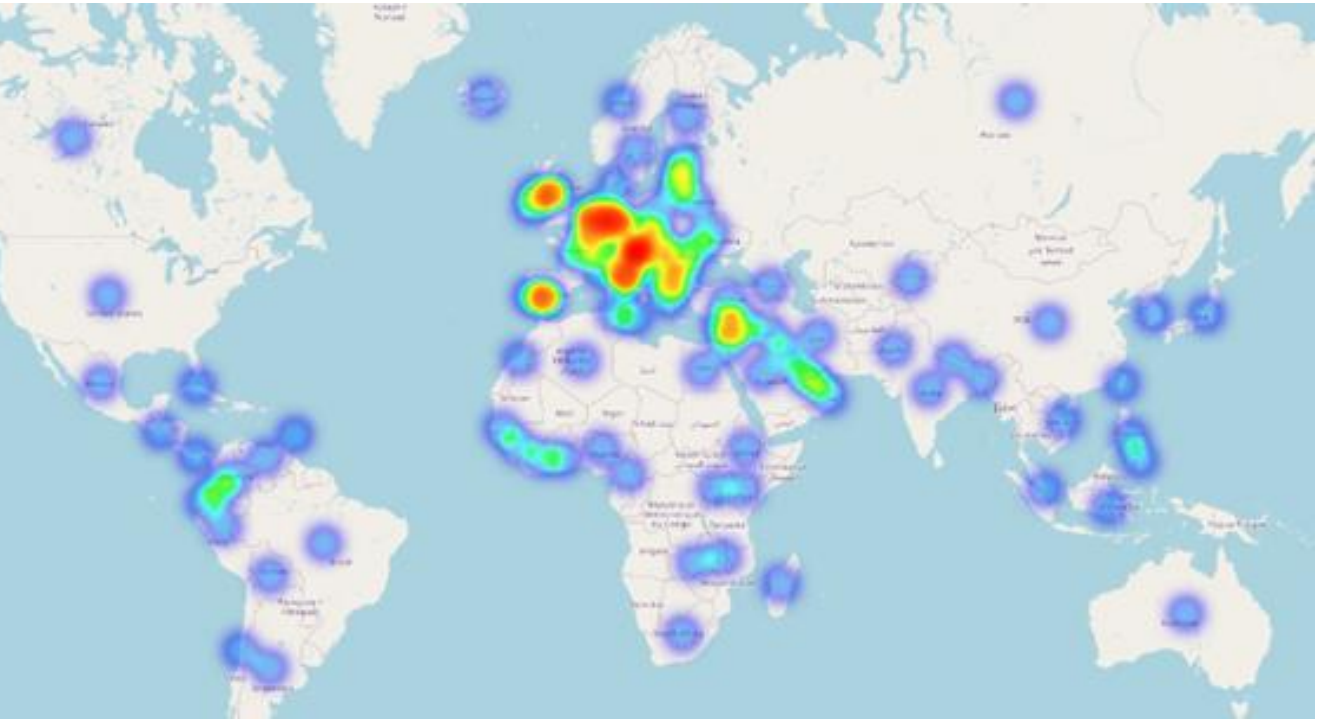


Figure 1. *Heatmap of global geographical distribution of data sources registered in the Catalogues.*

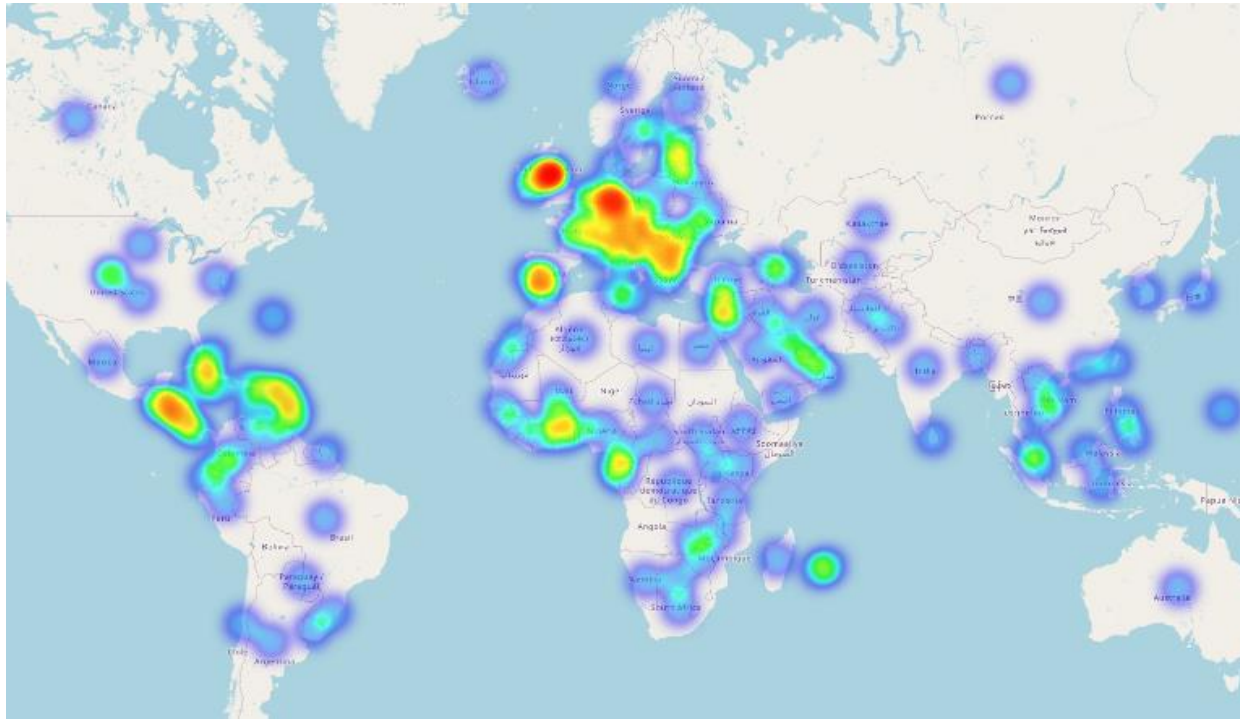


Figure 2. *Heatmap of global geographical distribution of studies registered in the Catalogues.*

List of metadata | Data sources

EUROPEAN MEDICINES AGENCY

I	II	III	IV	V
Administrative details	**Data elements collected**	**Vocabularies**	**Data flows and management**	**Quantitative descriptors**
• Data source ID • Name and acronym • Data holder (institutions look-up) • Data source contact • Countries & regions • Language • First established date • Time span: first collection – last collection • Website • Publications • Studies conducted using data source • Qualifications • Main financial support (European public funding, industry funding, public-private partnership etc.) • Data source type (e.g., biobank, cancer registry, hospital inpatient..) • Care setting (e.g. GP, community pharmacist, hospital outpatient..)	• **Exposure**: availability of data on prescriptions and/or dispensing, ATMPs, contraception, vaccines, other injectables, medical devices, procedures, medicinal products and indication, **genetic data** • **Outcomes**: availability of data on hospital admission or discharge, ICU admission, cause of death, clinical measurements, genetic data, patient-generated data, health care utilisation, diagnostic codes, specific diseases, with disease information collected • Disease information • Medicinal product information • Family linkage • Sociodemographic information • Lifestyle factors information	• Medicinal product vocabulary • Cause of death vocabulary • Quality of life measurement • Prescription vocabulary • Dispensing vocabulary • Indication vocabulary • Procedures vocabulary • Genetic data vocabulary • Biomarker data vocabulary • Diagnosis/medical event vocabulary	• Governance details: Documents or webpages that describe the overall governance, data capture and management (..) • Biospecimen access • Access to subject details • Description of data collection • Event triggered creation of record • Event triggering registration and de-registration of a person • Linked data sources, names and linkage variable. • Data management including possibility for data validation, data source preservation and approval for publication. • Informed consent for use of data research • Data source refresh and date • CDM mapping & name • ETL process and status	• Population age groups • Estimated percentage of population covered in catchment areas • Description of population in catchment area • Population size (by age groups) • Active population size (by age groups) • Median observation time between first and last available records for all and active individuals

List of metadata | Studies

EUROPEAN MEDICINES AGENCY

I

Administrative details

- EU PAS Register number
- Official study title and acronym
- Study description
- DARWIN EU study: yes/no
- Institution conducting study
- Network conducting study
- Study contact
- Primary lead investigator
- Study timelines
  - Contract signed
  - Study start
  - Data analysis start
  - Interim report
  - Final study report
- Study countries
- Source of funding
- Protocol doc
- Required by EU regulator Yes/No
- Required by RMP?
- Regulatory procedure number
- Other study ID

II

Methodological aspects

- Study topic (e.g. disease/health condition, herbal medicine, medical device, medical product/procedure etc.)
- Study type
- Name of medicine (brand name)
- Study drug INN
- ATC Code
- Medical condition
- Population studied
- Age groups
- Special population of interest (e.g. frail, nursing women, pregnant women, hepatic impaired etc.)
- Estimate number of subjects
- Study design
- Study objective
- Setting (persons, place, time period, selection criteria)
- Comparators
- Outcomes
- Data analysis plan (e.g. risk estimation, internal/external validity etc.)
- Summary results incl. study report
- Study publications

III

Data management

- ENCePP CoC Yes/No/NA
- ENCePP seal incl. documents
- Data sources (look-up)
- Data source types
- CDM mapping
- Check conformance
- Check completeness
- Check stability
- Check logical consistency
- Data characterisation conducted
- Procedure of data extraction
- Procedure of result generation

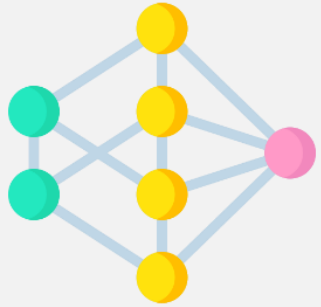
Stakeholder benefits | Overview

EUROPEAN MEDICINES AGENCY

Enhanced Data Visibility and Accessibility

- **Filter**, **sort** and **export** search results and records (Excel, pdf)
- Streamlined **identification of data sources** for specific research questions
- **Link** between **data sources** and **associated studies**
- **Easily accessible** on the EMA public website (pending implementation)

Continuous Improvement and Interoperability



- **Planned ongoing enhancements** and technical improvements (e.g. visualisations, dashboards)
- Data **interoperability and integration with other catalogues**
- Collaboration with future **EHDS EU datasets catalogue**

Transparency and Collaboration

- **Public access** to study reports and results
- **Encouraged collaboration** among different actors
- Towards **better-informed discussions** on drug safety and effectiveness
- Strengthened **collective effort** to advance RWD utilisation and RWE generation

A user would like **to identify suitable data source(s) for a planned study on pharmacogenomics**	The data sources Catalogue offers information on the **data source content** (e.g.: **genetic data collection**, disease, demographics), governance, accessibility, contact points to help the choice of data source. It allows benchmarking of different data sources referring to similar population when planning a study.
A user would like to **write a study protocol** or report for which they **need to describe a data source or wants to compare the characteristics of a specific data source** with other ones covering fully or partially the same population	The user can extract from the Catalogues standardised information on each data source and provide a reference (PURI or url) to public information for each of them. The Catalogues provide a harmonized description of the data source characteristics that allow comparison, information on common variables and information on possible linkages with other data sources allowing to harmonise data on the same individuals and provide additional information (e.g., on confounding).
A study protocol submitted uses a data source. A regulator needs to **understand the suitability of the data source** proposed	The study can be retrieved using the studies catalogue; the protocol is available. Other **similar studies** can be retrieved using studies structured metadata, and a comparison of **data sources** used in similar research is possible.
A user **reads a study report** for which they need to **evaluate the data source(s) used** in the study	The study report is available in the catalogue, along with details on the data source used. **Other studies** conducted using this data source can be consulted using the catalogues and provide orientation on the suitability of the data. The information on proposed data source used can be easily retrieved and assessed in the same context.

Resources & useful links

EUROPEAN MEDICINES AGENCY

HMA-EMA Catalogues of real-world data sources and studies
https://catalogues.ema.europa.eu/

Good Practice Guide (GPG)
https://www.ema.europa.eu/en/documents/regulatory-procedural-guideline/good-practice-guide-use-metadata-catalogue-real-world-data-sources_en.pdf

List of metadata for the HMA-EMA Catalogues of real-world data sources and studies
https://www.ema.europa.eu/en/documents/other/list-metadata-real-world-data-catalogues_en.pdf

ENCePP
https://encepp.europa.eu/index_en

Big Data Highlights
https://ec.europa.eu/newsroom/ema/newsletter-archives/54268

https://ec.europa.eu/newsroom/ema/items/835719/en

Multi-stakeholder webinar (4 March 2024)
https://www.ema.europa.eu/en/events/multi-stakeholder-webinar-hma-ema-catalogues-real-world-data-sources-studies

Workplan 2023-2025: HMA-EMA joint Big Data Steering Group
https://www.ema.europa.eu/en/documents/work-programme/workplan-2023-2025-hma-ema-joint-big-data-steering-group_en.pdf

Big Data
https://www.ema.europa.eu/en/about-us/how-we-work/big-data

Demonstration of the platform

Thank you!

Any questions?

Further information

RWDcatalogues@ema.europa.eu

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