8 April 2025
EMA/124437/20255
European Medicines Agency

User guide of the HMA-EMA Catalogues of real-world data sources and studies

Version 1.0

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

Contents

1. Purpose of this document

The User guide of the HMA-EMA Catalogues of real-world data sources (RWD) and studies has been developed for users submitting data to the Catalogues (i.e.: data holders and investigators submitting study data) to help them navigate the Catalogues in this context. The user guide provides descriptions of the data fields and definitions, as well as guidance on how to submit and maintain a record in the Catalogues.

A Good Practice Guide for the use of the HMA-EMA Catalogues of RWD sources and studies has also been published to provide regulators, researchers (including academia and pharmaceutical industry) and other interested stakeholders with recommendations on the use of the catalogues from the perspective of the data user. Specifically, the guide focuses on describing the steps and best practices on identifying a suitable data source, when planning a study.

2. Catalogue of real-world data sources: description of the data fields and definitions

Following several prioritisation exercises and consultations with stakeholders, the below metadata elements, which have been published by HMA/EMA in the List of metadata for the HMA-EMA Catalogues of real-world data sources and studies[1], have been selected for a first iteration of this process. The data elements aim to describe the data sources, with a view of facilitating the choice of data source for the specific use cases listed in the 'Good Practice Guide for the use of the HMA-EMA Catalogues of real-world data'.

2.1. Data elements characterising the data source

A data source is described by the data holder that sustains the collection of records in the data source, the underlying population that can potentially contribute records to a data source, and the prompt that leads to creation of a record in the data source.

2.1.1. Data source – Administrative details

2.1.1.1. Data source domain (C1.15)

The domain that the data pertains to, which can be either human or veterinary.

2.1.1.2. Species (C1.16)

This field is applicable only when the Veterinary domain is selected in *C1.15*. It specifies the species included in the data source. The look-up offers values synchronized with the data provided by the RMS service, where a centralisation across EMA databases is performed, for the purpose of interoperability and ease of searchability. More than one value can be selected if needed.

2.1.1.3. Name of the data source (C1.2)

The name of the data source, as used in European projects, must be provided. If the database is widely known by several names, these can be provided in this field, separated by a '/' sign. Where the data source has been known by different names in the past, these can be provided, using parenthesis with the note 'formerly known as'. Where the name of the data source is in a local language, the English translation should also be provided, using parentheses.

[1] HMA/EMA. List of metadata for Real World Data catalogues (2024).

2.1.1.4. Data source acronym (C1.3)

Where the data source is generally known under a specific acronym, this should be provided.

2.1.1.5. Data holder (C4.1)

The data holder must be provided, selecting one of the existing entries from the 'institutions' available look-up. For the purpose of this Catalogue, a data holder is defined as an organisation that sustains the collection of records in a data source.

2.1.1.6. Data source contact details (M1.3, M1.6)

At least one contact name and email (main contact) should be provided for queries related to the data source. An alternate contact name and email may also be provided. Functional (organisation) contact emails may be provided. The contact details should be kept up to date and will be made public.

2.1.1.7. Data source countries (C1.5)

The country where the data originate from should be selected from the list of country codes (ISO 3166-1). Where needed, multiple countries can be selected.

2.1.1.8. Data source language(s) (C6.2)

The data source language should be specified using the appropriate ISO 639 code.

2.1.1.9. Data source regions (C1.5.1)

The geographical regions that the data source covers should be provided using region codes (ISO 3166-2). Multiple regions can be selected where required.

2.1.1.10. Date when the data source was first established (C4.5)

The date when the data source was first set-up (e.g. system go-live date where applicable). This date can be different from the 'first collection date' (C1.12).

2.1.1.11. First collection date (C1.12)

The date when data started to be collected or extracted, where extraction of data is part of the data collection process e.g. extraction of data from a linked data source.
It is expected that this information is populated only once, when the data source is first described (with the exception of error corrections from the initial submission).

2.1.1.12. Last collection date (C1.13)

Where applicable, the date when the data collection ended. This information should only be provided for data sources where the data collection has stopped permanently.

2.1.1.13. Data source website (C11.1)

Where such information is available, a link to the dedicated webpage describing the data source should be provided. The information listed would capture information such as data content, release notes etc.

2.1.1.14. Data source publications (C11.2)

A list of peer-reviewed papers (e.g. linked to journal-indexing repositories) or documents describing the data source (validation, data elements, representativity) or its use in pharmacoepidemiologic research.

2.1.1.15. Data source qualification (C3.1, C3.1.1)

If the data source has successfully undergone a formal qualification process (e.g., from the EMA, or ISO or other certifications), this should be described.

2.1.1.16. Main financial support (C4.6)

The source of finance for the data source in the last three years should be specified using the below categories:

- European public funding
- Funding by own institution
- Funding from industry or contract research organisation
- Funding from public-private partnership
- Funds from patients organisations, charity and foundations
- National, regional, or municipal public funding

Where the source of finance does not fit in the above categories, the value 'Other' can be used.

2.1.1.17. Data source type (C5.1, C5.1.1)

A data source may fit into one or more of the following categories:

Administrative
- population registry
- death registry
- registration with healthcare system
- exemptions from co-payment
- diagnostic tests or procedures reimbursement
- administrative healthcare records (e.g., claims)

Primary care
- primary care medical records
- pharmacy dispensing records

Secondary care
- hospital discharge records
- hospital inpatient records
- hospital outpatient visit records
- emergency care discharge records
- specialist ambulatory care records

Registries
- birth registry
- induced terminations registry
- congenital anomaly registry
- cancer registry
- disease registry
- vaccination registry
- drug registry

Other
- biobank
- spontaneous reports of suspected adverse drug reactions
- mobile health (mHealth)

If none of the listed categories apply to the data source, the value 'Other' should be used and the type should be described in the available free text field (C5.1.1).

2.1.1.18. Care setting for data source (C1.14)

Where the data source describes a care setting, this can be further characterised as:

- primary care – general practice (GP), community pharmacist level

- primary care – specialist level (e.g. paediatricians)
- secondary care – specialist level (ambulatory)
- hospital inpatient care
- hospital outpatient care

Note about 'secondary care – specialist level (ambulatory)' and 'hospital outpatient care': 'hospital outpatient care', although also ambulatory care, should be selected separately. If none of the listed categories apply to the data source, the value 'Other' should be used.

2.1.2. Data source – Data elements collected

2.1.2.1. Data source content characteristics

To characterise the content of the data source, the specific data elements should be selected as applicable.

Value (yes/no)	Description
Specific diseases (C1.10)	Data source collects information with a focus on specific diseases. This might be a patient registry or other similar initiatives. If specific diseases are captured, further information will be included as detailed section 2.1.2.2.
Rare diseases (C6.12)	The data source captures rare diseases, where the prevalence of the condition in the EU is less than 5 in 10,000 people.
Pregnancy and neonates (C1.9)	Where data on pregnant women and neonates (under 28 days of age), infant, and child development.
Hospital admission discharge (C6.10)	Information on hospital admission and/or hospital discharge is available in the data source.
ICU admission (C6.10.1)	Information on intensive care admission available.
Cause of death (C6.11)	The cause of death is captured, either as structured or unstructured information.
Prescriptions of medicines (C6.13)	The data source contains information on prescriptions of medicines.
Dispensing of medicines (C6.14)	The data source contains information on dispensing of medicines.
Advance therapy medicinal products (ATMP) (C6.16)	Any information on use of a medicinal product for human use that is either a gene therapy medicinal product, a somatic cell therapy product, or a tissue engineered product as defined in Regulation (EC) No 1394/2007 [Reg (EC) No 1394/2007 Art 1(1)].
Contraception (C6.17)	Any information on use of any type of contraception (oral, injectable, devices etc.)
Indication for use (C6.18)	Therapeutic indication for the use of medicinal product. The aim is to capture whether the therapeutic indication for the use of the medicinal product is clearly recorded. It should not be answered as 'Yes' if it can be derived by proxies.
Medical devices (C6.20)	Where data source captures information on medicinal devices (e.g.: pens, syringes, inhalers).
Administration of vaccines (C6.19)	Information on any vaccines administered.

Value (yes/no)	Description
Procedures (C6.21)	Medical procedures (e.g. surgical interventions, tests).
Clinical measurements (C6.23)	Information on clinical measurements (e.g.: BMI, blood pressure, height).
Healthcare provider (C6.24.1)	Data on individual health professionals or a health facility organization licensed to provide health care diagnosis and treatment services including medication, surgery and medical devices.
Genetic data (C6.25)	Data related to genotyping, genome sequencing.
Biomarker data (C6.26)	The term “biomarker” refers to a broad subcategory of medical signs – that is, objective indications of medical state observed from outside the patient – which can be measured accurately and reproducibly. For example, haematological assays, infectious disease markers or metabolomic biomarkers.
Patient-reported outcomes (C6.31)	Patient-reported outcomes (e.g. quality of life).
Patient-generated data (C6.27)	Health-related data created, recorded, or gathered by or from patients (or family members or other caregivers) to help address a health concern.
Units of healthcare utilisation (C6.29)	Quantification of the use of services for the purpose of preventing or curing health problems (e.g.: number of visits to GP per year, number of hospital days).
Unique identifier for persons (C6.4)	Where applicable, if patients are uniquely identified.
Diagnostic codes (C6.9)	If diagnostic codes are captured, further information will be captured in section 5.1.5.11.
Medicinal product information (C6.15.3)	If medicinal product information are captured, further information will be included as detailed in sections 2.1.2.3. and 2.1.2.12.
Quality of life measurements (C6.15.2)	If quality of life measurements are captured, further information will be included as detailed in section 2.1.2.13.
Lifestyle factors (C6.8.1)	If medicinal product information are captured, further information will be included as detailed in section 2.1.2.14.
Sociodemographic information (C6.1.1)	If medicinal product information are captured, further information will be included as detailed in section 2.1.2.15.

2.1.2.2. Disease information collected (C1.10.1)

Where the data source collects information with a focus on specific diseases (e.g. patient registry or other similar initiatives), the disease or diseases for which information is collected should be specified in this field, using MedDRA terminology.

2.1.2.3. Medicinal product vocabulary used (C6.15.1)

Vocabulary	Description
AIC	Autorizzazione all'Immissione in Commercio ("The AIC number is a unique identifier assigned to each medicinal product that has received marketing authorization in Italy." Further reference here.
ART 57	Authorised medicines information in EU and EEA. Further reference here.
ATC	Anatomical Therapeutic Chemical (ATC) classification system - active substances divided into different groups according to the organ or system on which they act and their therapeutic, pharmacological and chemical properties. Further reference here
CNF	Classificazione Nazionale dei Farmaci, Italian classification system used by the Italian Agency (AIFA). Further reference here.
dm+d	Dictionary of descriptions and codes which represent medicines and devices in use across the NHS. Further reference here.
EQDM	European Directorate for the Quality of Medicines. Further reference here.
Gemscript	Gemscript is an integrated DM+D drug dictionary, adheres to the NHS standards to provide information to the specific medicines and devices used in the diagnosis or treatment of patients.
GTIN	Global Trade Item Number (GTIN) is used to identify trade items uniquely and unambiguously
IFA GmbH	Informationsstelle für Arzneispezialitäten. Further reference here.
MTHSPL	FDA Structured Product labelling. Further reference here.
NDC	FDA's National Drug Code (NDC) Directory contains information about finished drug products, unfinished drugs and compounded drug products. Further reference here.
NDF	National Drug Formulary – drug dictionary used by Singapore Ministry of Health. Further reference here
RxNorm	A normalized naming system for generic and branded drugs. Further reference here.
SNOMED	Terminology created and maintained by SNOMED International. Further reference here.
SPN	Standard Product Nomenclature. Further reference here.
WHO Drug	Drug dictionary created and maintained by WHO. Further reference here.
Z-index (G-standard)	External reference here.

Where the medicinal product information is not coded (i.e.: provided as free text) this should be marked accordingly (i.e.: Not coded/Free text).

If other dictionaries than the listed ones are used, the value 'Other' should be used.

2.1.2.4. Cause of death vocabulary (C6.11.1, C6.11.2)

Vocabulary	Description
CCS	Vocabulary used for the Clinical Classifications Software. Further reference here.
CPT	Current procedural terminology is a standard vocabulary for surgical procedures, minor procedures that physicians perform in the office, radiology tests, and a small number of laboratory tests. Further reference here.
dm+d	Dictionary of descriptions and codes which represent medicines and devices in use across the NHS. Further reference here.
HCPCS	Healthcare Common Procedure Coding System. External reference here.
Human Phenotype Ontology (HPO)	External reference here.
ICD	International Classification of Diseases. External reference here.
ICD-10	International Classification of Diseases, 10th revision. External reference here.
ICD-10-CM	International Classification of Diseases, 10th revision, Clinical Modification. External reference here.
ICD-11	International Classification of Diseases, 11th version. External reference here.
ICD-9	International Classification of Diseases, 9th revision. External reference here.
ICD-9-CM	International Classification of Diseases, 9th revision, Clinical Modification. External reference here.
ICPC	International Classification of Primary Care. Further reference here.
ICPC-1	International Classification of Primary Care, version 1.0. External reference here.
ICPC-2	International Classification of Primary Care, version 7.0. External reference here.
MedDRA	Medical Dictionary for Regulatory Activities. Further reference here.
OMIM	Online Mendelian Inheritance in Man. External reference here.
OPCS	Classification of Interventions and Procedures. Further reference here.

Vocabulary	Description
OPS	Operation and procedure classification system. External reference here.
Orphacode	Orphanet rare disease nomenclature. External reference here.
Orphanet Rare Disease Ontology (ORDO)	External reference here.
Read	External reference here.
SNOMED	Systematized Nomenclature of Medicine. Further reference here.
SNOMED CT	Systemized Nomenclature of Medicine – Clinical Terms. Further reference here.

Where the cause of death is not coded (i.e.: provided as free text) this should be marked accordingly.

If other dictionaries than the ones listed are used, the value 'Other' should be used.

2.1.2.5. Prescription vocabulary (C6.13.1)

Vocabulary	Description
ATC	Anatomical Therapeutic Chemical (ATC) classification system - active substances divided into different groups according to the organ or system on which they act and their therapeutic, pharmacological and chemical properties. Further reference here.
DrugBank	DrugBank Online is a comprehensive, free-to-access, online database containing information on drugs and drug targets. Further reference here.
EphMRA	Anatomical Classification of Pharmaceutical Products maintained by EphMRA. Further reference here.
RxNorm	A normalized naming system for generic and branded drugs. Further reference here.

Where the prescription is captured but not coded (i.e.: provided as free text) this should be marked accordingly.

If other dictionaries than the ones listed are used, the value 'Other' should be used. In this case, the prescription vocabulary used should be provided in the free text field accordingly.

2.1.2.6. Dispensing vocabulary (C6.14.1)

The dictionary used to code the dispensing information captured in the data source should be selected.

For further details on the available values, see section 2.1.2.5.

Where the dispensing information is captured but not coded (i.e.: provided as free text) this should be marked accordingly.

If other dictionaries than the ones listed are used, the value 'Other' should be used. In this case, the dispensing vocabulary used should be provided in the free text field accordingly.

2.1.2.7. Indication vocabulary (C6.18.1, C6.18.2)

The dictionary used to code the therapeutic indication captured in the data source should be selected. For further details on the available values, see section 2.1.2.4.

Where the therapeutic indication is captured but not coded (i.e.: provided as free text) this should be marked accordingly.

If other dictionaries than the ones listed are used, the value 'Other' should be used. In this case, the indication vocabulary used should be provided in the free text field accordingly (C6.18.2)

2.1.2.8. Procedures vocabulary (C6.22)

The dictionary used to code the procedures captured in the data source should be selected.
For further details on the available values, see section 2.1.2.4.

Where the procedure is captured but not coded (i.e.: provided as free text) this should be marked accordingly.

If other dictionaries than the listed ones are used, the value 'Other' should be used. In this case, the procedures vocabulary used should be provided in the free text field accordingly.

2.1.2.9. Genetic data vocabulary (C6.25.1)

Vocabulary	Description
EGO	Eukaryotic Gene Orthologues. Further reference here.
GO	Gene Ontology. Further reference here.
HGNC	HUGO Gene Nomenclature Committee. External reference here.
HGVS	Human Genome Variation Societ. External reference here.
OGG	A biological ontology in the area of genes and genomes. Further reference here.
OMIM	Online Mendelian Inheritance in Man. External reference here.
PHARE	PHArmacogenomic RElationships Ontology. Further reference here.
SOPHARM	Suggested Ontology for Pharmacogenomics. Integrates OBO ontologies and formalizes specific gene variants. Further reference here.

Where the genetic data is captured but not coded (i.e.: provided as free text) this should be marked accordingly.

If other dictionaries than the listed ones are used, the value 'Other' should be used. In this case, the genetic data vocabulary used should be provided in the free text field accordingly.

2.1.2.10. Biomarker data vocabulary (C6.26.1)

Vocabulary	Description
FOBI	Food-Biomarker Ontology. Further reference here.
HPO	Human Phenotype Ontology. External reference here.
SMASH	Semantic Mining of Activity, Social, and Health data. Further reference here.

Where the biomarker data is captured but not coded (i.e.: provided as free text) this should be marked accordingly.

If other dictionaries than the listed ones are used, the value 'Other' should be used. In this case, the biomarker vocabulary used should be provided in the free text field accordingly.

2.1.2.11. Diagnosis/ medical event vocabulary (C6.9.1)

The dictionary used to code the diagnosis, or any other medical event captured in the data source should be selected.

For further details on the available values, see section 5.1.5.3.

Where the diagnosis or medical event is captured but not coded (i.e.: provided as free text) this should be marked accordingly.

If other dictionaries than the listed ones are used, the value 'Other' should be used. In this case, the diagnosis/ medical event vocabulary used should be provided in the free text field accordingly.

2.1.2.12. Medicinal product information available (C6.15)

The type of information captured with regards to the medicinal product should be selected from the values described in the table below.

Vocabulary	Description
Active ingredient(s)	The active ingredient(s) of the medicinal product
Brand name	Specific name or trademark under which a medicine is sold
Batch number	The designation printed on the medicine label that allows the history of its production to be traced
Formulation	Pharmaceutical form of the medicinal product (e.g.: tablets, capsules etc.)
Strength	The amount of active ingredient contained in the medicinal product

Vocabulary	Description
Package size	Number of individual formulations contained in a package (e.g.: 30 tablets per package)
Dose	The medicinal product dose prescribed or administered to the patient
Dosage regime	The schedule of doses of a medicinal product per unit of time (e.g.: every 6 hours)
Route of administration	The manner in which a medicinal product enters the body (e.g.: oral, intravenous)

2.1.2.13. Quality of life measurements (C6.28, C6.28.1)

Vocabulary	Description
AQoL-8D	Assessment of Quality of Life 8-Dimension. Further reference here.
QOLS	Quality of Life Scale. Further reference here
MQOL	McGill Quality of Life Questionnaire. Further reference here.
MQOL-E	The McGill Quality of Life Questionnaire – Expanded. Further reference here.
HRQOL	Health-related quality of life. Further reference here.
WHOQOL	World Health Organization External Measuring Quality of Life. Further reference here.
EQ5D	Standardised measure of health-related quality of life developed by the EuroQol Group. EQ-5D assesses health status in terms of five dimensions of health. Further reference here.
15D	The 15D is a generic 15-dimensional self-administered instrument for measuring HRQoL (Health-related quality of life). Further reference here.
SF-12	
SF-36	The Short Form (36) Health Survey is a 36-item, patient-reported survey of patient health. Further reference here.
SF-6D	An abbreviated variant of SF-36 commonly used in health economics as a variable in the quality-adjusted life year calculation to determine the cost-effectiveness of a health treatment. External reference here.
HUI	Health Utilities Index. Further reference here.

Where the quality of life is captured but not coded (i.e.: provided as free text) this should be marked accordingly.

If other dictionaries than the ones listed are used, the value ‘Other’ should be used. In this case, the name of the ‘quality of life’ scale used should be provided in the free text field accordingly (C6.28.1).

2.1.2.14. Lifestyle factors (C6.8)

Where the data source captures this information, one or more of the following lifestyle factors can be selected:

- tobacco use
- alcohol use
- amount of physical exercise
- diet

2.1.2.15. Sociodemographic information collected (C6.7)

Where one or more of the following specific sociodemographic information are captured by the data source, these should be selected:

- age
- gender
- ethnicity
- country of origin
- indicator of socioeconomic status
- marital status
- education level
- type of residency
- living in rural area
- health area
- deprivation index

2.1.3. Data source – Quantitative descriptors

This section aims to collect a limited amount of data elements that look at population qualitative details and quantitative details of the data source.

2.1.3.1. Population age groups (C1.8)

The information on the following age groups are being captured separately:

- in utero;
- paediatric population (< 18 years):
- neonate:
  - preterm newborn infants (0 - 27 days);
  - term newborn infants (0 - 27 days);
  - infants and toddlers (28 days - 23 months);
  - children (2 to < 12 years);
  - adolescents (12 to < 18 years);
- adult and elderly population (>18 years):
  - adults (18 to < 65 years):
    - adults (18 to < 46 years);
    - adults (46 to < 65 years);
  - elderly (≥ 65 years):
    - adults (65 to < 75 years);
    - adults (75 to < 85 years);
    - adults (85 years and over).

2.1.3.2. Population covered by the data source (C1.11.2)

The percentage of the population covered by the data source in the catchment area should be specified.

2.1.3.3. Population not covered by the data source (C1.11.1)

The description of the population covered by the data source in the catchment area whose data are not collected, where applicable (e.g.: people who are registered only for private care).

2.1.3.4. Family linkage (C6.6, C6.6.1)

Where family linkage is made available in the data source this should be characterised using one or more of the following values (C6.6.1): household (where the information on individuals sharing a household can be identified), mother-child, father-child, sibling (maternal or paternal siblings).

If family linkage is not available, it should be specified if familial linkage can be created on an ad-hoc basis (C6.6).

2.1.3.5. Population size (C7.1)

The total number of unique individuals with records captured in the data source.

2.1.3.6. Population size by age (C7.3)

Where this information can be extracted, the number of unique individuals split by age groups should be captured.

For individuals who are not lost to follow-up, their current age at the time of filling out the webform should be used to determine the appropriate age group. If there is uncertainty about whether an individual is lost to follow-up, their age should be based on the date of their last observation in the data source.

2.1.3.7. Active population size (C7.1.1)

Number of individuals alive and currently registered with active records in a data source. For example, an active population for administrative healthcare data refers to the collection of patients for which there is an active record, i.e. the record was created and not closed because the patient moved or died.

2.1.3.8. Active population size by age (C7.3.1)

Where this information can be extracted, the number of unique active individuals split by age groups should be captured.

For example, an active population for administrative healthcare data refers to the collection of patients for which there is an active record, i.e. the record was created and not closed because the patient moved or died.

2.1.3.9. Median observation time (B6.3)

The median time, in years, between first and last available records for unique individuals captured in the data source.

2.1.3.10. Median observation time active (B6.3.1)

The median time, in years, between first and last available records (e.g.: due to death, loss to follow-up, end of follow-up) for unique *active* individuals (alive and currently registered) captured in the data source.

For example, an active population for administrative healthcare data refers to the collection of patients for which there is an active record, i.e. the record was created and not closed because the patient moved or died.

2.1.4. Data source – Data flows and management

2.1.4.1. Governance details (C2.3)

Description of the documents or links to webpages that describe the overall governance, processes and procedures for data capture and management, data access including limitation(s) in the use of the data due to privacy concerns, data quality check and validation results, utilisation for research purposes.

2.1.4.2. Follow-up (C2.13, C2.13.1, C2.7)

If further follow-up would be needed, the availability of below access options should be specified:

- Accessing biospecimens: if this is possible (C.2.13) then also the biospecimen access conditions should be described (or a reference source can be added) (C2.13.1)
- Contacting patients or practitioners (C2.7)

2.1.4.3. The process of collection and recording (C4.3)

The process or manner in which recording of data in the data source occurs should be described; this could include the tools used, such as surveys, or a description of the system that the data holder uses to gather data and store the data source.

2.1.4.4. Record creation (C5.2)

An event occurring that triggers the creation of a new record in the data source should be described (e.g.: hospital discharge, specialist encounter, medicinal product dispensing).

This refers to the creation of a record in the data source (and not to the registration of a person, see below).

2.1.4.5. Registration of a person (C1.6, C1.6.1)

An event occurring that triggers the registration of a new person in the data source should be selected from the following available values:

- Birth
- Immigration
- Residency obtained
- Start of insurance coverage
- Disease diagnosis
- Start of treatment
- Practice registration

Where none of the above values apply, the triggering event for a person to be registered in the data source should be described separately (C1.6.1).

2.1.4.6. De-registration of a person (C1.7, C1.7.1)

The event triggering de-registration of a person in the data source. The event triggering de-registration of a person in the data source should be selected from the following available values:

- Death
- Emigration
- End of insurance coverage

- Practice deregistration
- Loss to follow up
- End of treatment

Where none of the above values apply, the triggering event for a person to be de-registered in the data source should be described separately (C1.7.1). The triggering event should be captured in the data source.

2.1.4.7. Linkage (B5.2, B5.2.1, B5.3, B4.1)

Where the data source is created by the linkage of other data sources or if linkages to other data sources are possible on an ad-hoc basis, the elements of the linkage should be briefly captured as follows:

- The linkage strategy (B5.2): whether the linkage is deterministic, probabilistic or a combination of the two.
- The linkage variable used (B5.2.1) (e.g.: patient ID, date of birth etc.)
- The completeness of the linkage (B5.3), described as a percentage along with the reference used
- Names of the linked data sources (B4.1). Where these data sources are available in the data source Catalogue, these should be cross-referenced.

2.1.4.8. Data management specifications (C2.7, C8.5, C8.5.1, C2.9):

The following information related to data management specifications should be selected, as applicable to the data source:

- Whether or not the data source allows data validation (e.g.: access to original medical charts)
- If the records are preserved indefinitely (C8.5);
- Where the records are not indefinitely preserved, the number of years for which the records are kept should be specified (C8.5.1)
- Whether approval is needed for publishing results of a study using its data (C2.9)

2.1.4.9. Informed consent for use of data for research (C2.5, C2.5.1)

The need for informed consent in the context of research should be captured here. The type of informed consent could be categorised as:

- Not required
- Required for general use of the data source
- Required for all studies run on the data source
- Required for intervention studies only
- Waiver

Where the informed consent does not fit in the above categories, the value ‘Other’ can be used and further details should be provided (C2.5.1).

2.1.4.10. Data source refresh (C8.2)

Where the data source is refreshed on fixed dates around the year, this should be provided by selecting the month as applicable (e.g.: every June). The field can be repeated where the refresh happens more often than once a year (e.g.: every May and November).

2.1.4.11. Data source last refresh (C8.3)

Where the data source is refreshed at particular times throughout the year, the date when the last refresh of the data source occurred should be provided.

2.1.4.12. CDM (Common Data Model) specifications (D1.2.1.1, D1.2, D1.2.1, D1.4, D1.7)

The following data elements should be captured for data sources being transformed using a Common Data Model (CDM), (D1.2.1.1) as follows:

- The CDM name should be selected from the existing predefined list as follows (D1.2):

Common Data Model	Description
BIFAP	Further reference here.
CDISC SDTM	Further reference here.
ConcepTION CDM	Further reference here.
CTcue Datamodel	Further reference here.
EUROCAT	Further reference here.
i2b2	Further reference here.
NorPreSS	A common data model developed for the Nordic Pregnancy Drug Safety Studies. Further reference here.
OMOP	The observational Medical Outcomes Partnership Common Data Model. Further reference here.
PCORnet	The National Patient-Centred Clinical Research Network Common Data Model. Further reference here.
PEDSnet	Further reference here.
Sentinel	Sentinel Common Data Model (SCDM) lead by the Sentinel Operations Center (SOC) residing within the Harvard Pilgrim Health Care institute. Further reference here.
Vaccine Safety Datalink (VSD) Data Dictionary	Further reference here.
TrineTX	Further reference here.
EUROMEDICAT	Further reference here.

Where the CDM used is not listed in the values offered, further details should be provided (D1.2.1)

- The CDM website reference should be provided where available (D1.4)
- The CDM release frequency, in number of months, should be provided (D1.7)

2.1.4.13. Data source ETL (Extract, Transform, Load) to a CDM (Common Data Model) (B7.1, B7.5, B7.3, B7.4)

Where applicable, further information on the ETL to a CDM should be provided as follows:

- The status of the ETL of the data source should be described as either: planned, in progress or completed.
- The frequency of the ETL in months
- The version(s) of CDM(s) to which the data source has been ETL-ed
- Data source ETL specifications: documents describing the mapping of the data source to the CDM (including codes and scripts to transform original data to CDM)

3. Registering a Data source

Any data holder may submit an entry in the HMA-EMA Catalogue of real-world data sources, regardless of the country of origin.

The questionnaire consists of 16 questions and are divided into four steps:

1. Administrative details,
2. Data elements collected,
3. Quantitative descriptions, and
4. Data flows and management.

All mandatory fields, marked with a red asterisk (*), will need to be completed in order to move to the next step. A sample questionnaire for offline review can be found on the support page. A draft can only be saved once all the mandatory fields have been filled in. We strongly encourage to fill in as many fields as possible, and to be as descriptive and detailed as possible in the description fields. Refer to section 2. Catalogue of real-world data sources: description of the data fields and definitions for further explanations and definitions on the metadata fields. Multiple users may collaborate on a single data source record, as needed. Please see section 8.6. 8.6. Adding a co-author on how to submit a collaboration request.

Important to note is that in order to submit a data source, the relevant data holder of that data source must already be registered as an *institution* in the RWD Catalogues. To submit an institution, refer to section 6.

To enter a data source in the RWD Catalogue, follow the link: Catalogue of RWD sources | HMA-EMA Catalogues of real-world data sources and studies (europa.eu). Ensure that you are logged in, using your EU Login account credentials. If you do not have an EU Login account, visit the support page on how to create one.

3.1. Validation and publication process

EMA is responsible for validating the content that is submitted for publication in the Catalogue of real-world data sources. The EMA Validator uses the following criteria to assess the submission:

- Relevance to Health Context, with a focus on data sources supporting medicines regulation. The data source should be considered relevant to the healthcare domain, aligning with EMA's regulatory focus and objectives. As one of the main objectives of the Catalogue of RWD sources is to promote data discoverability, this assessment is made in an inclusive manner, aiming to reflect the diversity of existing landscape of data sources.
- Uniqueness and Non-Duplication: A check of the data source as a 'new' entity is performed in the context of the other data sources already registered in the Catalogue. An assessment of a potential record duplication is carried out and clarifications are sought from the Editor(s) when record duplication is suspected. Data source records created through linkage of other data sources, as well as data sources resulting from the transformation process connected to a common data model are considered independent (unique) data entries, while the connections between these data sources should be documented in the metadata elements. The criteria describing what constitutes a duplicate record are adjusted and further aligned as further information is received via communication with the Editor(s).
- Comprehensiveness and inclusion of mandatory elements: The data source record

should contain at a minimum the complete set of mandatory metadata elements – the provision of this information is enforced technically; the EMA Validator performs an initial content validation of the mandatory fields, ensuring that the content is meaningful and correct, to the extent that correctness can be assessed.

If the submission does not meet all the validation criteria described, the EMA Validator will return the submission using the Catalogues' data management notification system, adding a message in the Revision Log containing the justification for the return. Subsequently, the Editor may further refine the submission with additional information (as required) and resubmit. This process is repeated until all the criteria have been met.

Once all the criteria have been met, the EMA Validator will verify that the data source submission is *complete*, meaning that the metadata information expected on a given data source is submitted. This encompasses information in both mandatory fields and non-mandatory fields. The expected data may vary between data sources. For example, if the data source type is a 'disease registry', it may be expected that the information on a specific disease is captured and specified in the relevant fields (under data elements collected). Another example is if a data source type contains 'pharmacy dispensing records', the relevant dispensing vocabulary should be identified (under data elements collected). While these fields are not mandatory, EMA may return the submission if this information is incomplete/inaccurate.

At this stage, the accuracy of the submission is also verified. The EMA Validator verifies the information submitted, in its entirety, against publicly available information (e.g., data source website) and if it is accurate.

If the findings, related to the completeness and accuracy of the submission, do not pass the assessment, and cannot be resolved by minor corrections by the EMA Validator, the submission is returned to the Editor for necessary corrections and/or further details. It is recommended that the Editor provides the necessary changes and resubmits the entry within 2-3 weeks for validation. If the Editor does not update the entry, the entry will remain unpublished. The EMA Validator may perform minor corrections on the data source entry by correcting evident typos and/or harmonizing the entry. For example, if the Editor enters information in a free-text field (e.g. "we use MedDRA to code indication"), while a look-up field above is available to select 'MedDRA'. Any other inaccuracies will be considered major and will be returned to the Editor for correction.

Once the data source entry passes the EMA assessment, the EMA Validator approves the data source which triggers the automatic publication of the data source entry in the Catalogue of real-world data sources and the entry is now available for public viewing.

Please note, the Editor will receive email notifications from the following domain: fpfis.tech.ec.europa.eu. If you have not received a notification, please check your spam inbox or contact us.

3.2. Maintenance of information in the Data source Catalogue

It is important that the metadata information published in the Catalogue of real-world data sources is kept up to date. The Editor is expected to keep the information of their respective data source record(s) up to date on a regular basis. It is strongly encouraged to provide updates as often as needed, but at a minimum this is expected on a yearly basis. In special circumstances, for which data collection has concluded, the Editor may notify the EMA Validator through the revision log of the Catalogues maintenance system that the data collection for a particular data source has been finalised and no further updates are foreseen/necessary.

To provide an update, the Editor will need to login in order to access their dashboard which contains the list of entries associated to their account. The Editor can edit the entry at any time. Once the information

in the entry is updated, the entry will need to be submitted for validation once more before the update can be published. The EMA Validator will carefully review and assess the updated information according to the criteria and steps outlined in section 3.1. Validation and publication process. If the Editor does not provide an update of the data source entry within two years since the last update, the data source will be marked as “outdated”. The entry will remain public to support assessment of study data and data discoverability. The Editory may update the data source record at any point after the two-year period after which the outdated flag will be automatically removed.

Alternatively, the Editor may request the data source entry to be unpublished. For example, if a data source ceases to exist or is no longer in use. The Editor will need to send a request via the service desk to the EMA Validator to unpublish the data source entry, providing sufficient justification for this request. The EMA Validator will carefully consider the request and justification. The timelines to process this request may take up to two weeks or up to one month in exceptional circumstances.

4. Catalogue of real-world data studies: Description of the data fields and definitions

Following several prioritisation exercises and consultations with stakeholders, the below metadata elements, which have been published by HMA/EMA in the List of metadata for the HMA-EMA Catalogues of real-world data sources and studies, have been selected for a first iteration of this process. The data elements aim to describe the (observational) studies.

4.1. Data elements characterising the study

4.1.1. Study – Administrative details

The Catalogue of Real-World data studies defines three identifiers for each study; these identifiers are assigned automatically to a study the first time a study is created. The identifiers serve different purposes and are generated upon the first submission of a study.

4.1.1.1. PURI (Persistent URI)

A study identifier formatted as a link. The generation of a persistent URI is intended to support future exchange of information between HMA/EMA Catalogue and other catalogues.

The format of the identifier is as follows:

	Format
Study registered prior to February 2024	https://redirect.ema.europa.eu/resource/XXXXX or https://redirect.ema.europa.eu/resource/XXXX
Study registered after February 2024	https://redirect.ema.europa.eu/resource/XXXXXXXXXX

The numeric sequence matches the Study ID (see 3.1.1.3.)

4.1.1.2. EU PAS register number: (F2.2)

The EU PAS number is a study identifier assigned automatically to a study the first time is created. The format of this identifier is assigned as follows:

	Format	Example
Study registered prior to February 2024	EUPASXXXX or EUPASXXXXX	EUPAS2631 or EUPAS79097
Study registered after February 2024	EUPASXXXXXXXXXX	EUPAS1000000086

For studies registered after February 2024 the numeric sequence matches the Study ID.

4.1.1.3. Study ID

Study ID is a numeric study identifier of four or five digits for legacy studies registered in the former EU PAS register. For studies recorded for the first time in the HMA/EMA Catalogues an identifier of ten digits will be automatically created.

	Format	Example
Study registered prior to February 2024	XXXX or XXXXX	2631 or 79097
Study registered after February 2024	XXXXXXXXXX	1000000086

*4.1.1.4. Official study title and acronym * (F1.2)*

The official title of the study as found in the protocol, results or publications, where applicable.

The acronym, if one exists, should be added in parenthesis after the study title.

4.1.1.5. Study description (F10)

A brief, concise description of the key elements of a study (such as main objectives and where applicable, main results).

This field is limited to 2000 characters.

4.1.1.6. DARWIN EU® study (F9.5)

Studies developed as part of DARWIN EU ® are marked accordingly with the specific flag. This information is updated by the Coordination Centre/EMA.

Details on DARWIN EU® studies can be found here.

4.1.1.7. Study status (F11)

The Study Status is an automatically generated, calculated field based on the dates provided in fields F19 to F23 and F19.1 to F23.1, respectively – see section 3.1.1.16 for further information on providing this information. For studies that have either of the fields XX / YY checked, the values ‘Cancelled’ or ‘Discontinued’ are automatically applied.

The values generated in this field are as follows:

- ‘Planned’ when the planned or actual date of the ‘Date when funding contract was signed’ (F19/F19.1) are provided;

- ‘Ongoing’ when the actual date of ‘Study start date/protocol finalisation’ (F20.1) is provided;
- ‘Finalised’ when the actual date of ‘Date of final study report’ (F23.1) is provided;
- ‘Cancelled’ when a study was planned but cancelled before commencement;
- ‘Discontinued’ when a study was initiated but terminated before completion.

4.1.1.8. Institution conducting the study (F1.3)

The institution vocabulary of the Catalogue is managed as a vocabulary, striving for a standardisation of the terminology used. Where the institution required is not already present in the list, pre-registration of the institution entity is strongly encouraged, following the process described in Section 5 of this guidance.

It is alternatively possible to register a study using field F1.7 and update it at a later stage with the institution entity created.

4.1.1.9. Additional institutions if not in the list (F1.7)

This field can be used when an institution is not found in the look-up provided. It is strongly encouraged to register the institution entity separately using the process described in Section 5 of this guidance.

4.1.1.10. Network conducting the study (if applicable) (F1.8)

The network vocabulary of the Catalogue is managed as a vocabulary, striving for a standardisation of the terminology used. Where the network required is not already present in the list, pre-registration of the network entity is strongly encouraged, following the process described in Section 6 of this guidance.

It is alternatively possible to register a study using field F1. and update it at a later stage with the network entity created.

4.1.1.11. Additional networks if not in the list (F1.8.1)

This field can be used when a network is not found in the look-up provided. It is strongly encouraged to register the network entity separately using the process described in Section 6 of this guidance.

*4.1.1.12. Name of institution contact for study * (F1.4)*

The contact point for the study, on behalf of the institution running it, is provided here. In some cases, this might be the lead investigator.

*4.1.1.13. Study contact email * (F1.5)*

The e-mail address of the contact point provided in F1.4. The e-mail does not need to be linked to a personal address but can be a collective mailbox address or the address of another designated individual can be indicated for further correspondence (e.g. for receipt of acknowledgment emails).

This e-mail address will be made public.

4.1.1.14. Primary lead investigator name (F12)*

The name of the person assigned as the main point of contact for a registered study. In situations where no (primary) lead investigator has been nominated by the sponsor the contact details of the

person in charge of the conduct of the study should be entered in the respective fields. The staff member of a pharmaceutical company or institution who is entering the study details into the Catalogue can provide his/her contact details under the Contact point for institution (F1.4) if suitable.

4.1.1.15. Primary lead investigator ORCID (F1.6)

Where applicable, this field can be used to provide the ORCID (Open Researcher and Contributor ID) of the primary lead investigator.

ORCID is a unique, persistent identifier available to any researcher, widely used in the scientific committee; further information found here.

4.1.1.16. Study timelines: initial administrative steps, progress reports and final report

	Planned	Actual
Date when funding contract was signed *	(F19)	(F19.1)
Study start date *	(F20)	(F20.1)
Data analysis start date	(F21)	(F21.1)
Date of interim report, if expected	(F22)	(F22.1)
Date of final study report*	(F23)	(F23.1)

The dates provided in these fields are used to calculate the status of the study following the logic described in section 3.1.1.7.

The system mandates that at least one of the planned or actual date are entered for fields F19/F19.1, F20/F20.1 and F23/F23.1.

Date when funding contract was signed:

Date at which the contract between the funder of the study and the organisation in charge of conducting the study (or the primary/lead investigator) has been signed. In case of multiple funding contracts, the date of signature of the first contract relating to the study should be provided.

Study start date:

This date refers to the start of data collection. In case of primary data collection, this refers to the date from which information on the first study subject is first recorded in the study dataset. In the case of secondary use of data, the study start date is the date from which data extraction starts (NB: for multi-database studies, this is the date on which the first data extraction is performed).

Data analysis start date:

In line with the definition of the “end of data collection” provided in the guideline on Good Pharmacovigilance Practices (GVP) module VIII, chapter VIII.B.2, the date from which the analytical dataset is completely available also means “start date of data analysis”. The date for the end of data collection, which is required in the context of the submission of the final study report, should therefore be entered in the field “Data analysis start date”.

*4.1.1.17. Study countries in which this study is being conducted * (F1.9)*

The country (or countries) where the study is being conducted are mandatory information to be submitted here.

4.1.1.18. Source of funding (F8.8)

The source of funding can be selected from the following values:

- EMA
- EU institutional research programme
- National competent authority (NCAs)
- No external funding
- Non for-profit organisation (e.g., charity)
- Non-EU institutional research programme
- Other public funding (e.g.: hospital, university)
- Pharmaceutical company and other private sector

Where none of the above are applicable, the value 'other' can be selected. Further details can then be provided in field F8.8.1.

4.1.1.19. More details on source funding. (F8.8.1)

This field can be used to provide additional information on the source of funding (F8.8). If the source funding is a pharmaceutical company, the company name should be provided in this field.

*4.1.1.20. Study required by a regulator * (F14)*

This field should be flagged as 'Yes' if:

- the study is a PASS imposed as an obligation by a competent authority in accordance with Articles 10 or 10a of Regulation (EC) No 726/2004 or with Articles 21a or 22a of Directive 2001/83/EC, or

- the study is a PASS which is initiated, managed or financed voluntarily by a marketing authorisation holder and required in the Risk Management Plan (RMP) to further investigate safety concerns or

- any other study requested by a regulatory authority for any other reason.

*4.1.1.21. Is the study required by a Risk Management Plan (RMP) * (F14.1)*

- *EU RMP category 1 (imposed as condition of marketing authorisation)*

  A post-authorisation safety study (PASS) may be imposed as condition of the marketing authorisation because it is key to the benefit-risk profile of the product. In the EU Risk Management Plan (EU RMP) these studies are referred to as category 1 studies in the pharmacovigilance plan of an authorised medicinal product. If the condition is a non-interventional PASS, it will be subject to the supervision set out in Art 107 (m)-(q) of Directive 2001/83/EC and the format and content of such non-interventional PASS as described in Implementing Regulation 526/2012 Annex III (see GVP Module VIII)

- *EU RMP category 2 (specific obligation of marketing authorisation)*

  A PASS may be a specific obligation in the context of a conditional marketing authorisation (MA) or a MA under exceptional circumstances. In the EU RMP these studies are referred to as category 2 studies in the pharmacovigilance plan of an authorised medicinal product. If the specific obligation is a non-interventional PASS, it will be subject to the supervision set out in Article 107 (m)-(q) of Directive 2001/83/EC and the format and content of such non-interventional PASS as described in Implementing Regulation 526/2012 Annex III (see GVP Module VIII).

- *EU RMP category 3 (required)*

  PASS which do not fall in category 1 or 2 but are required to investigate a safety concern as part of the pharmacovigilance plan of an authorised medicinal product are legally enforceable. In the EU RMP these studies are referred to as category 3 studies (see GVP Module VIII).

- *Non-EU RMP only*

PASS which are included in risk management systems outside the jurisdiction of EU medicines regulation (e.g. Risk Evaluation and Mitigation Strategies (REMS) under US regulation).

- *Not applicable*

Any other study which is not subject to regulatory supervision and not a RMP pharmacovigilance activity.

4.1.1.22. Regulatory procedure number (RMP category 1 and 2 studies only) (F14.2)

This field is used to provide the regulatory authority's procedure number to facilitate compliance with regulatory requirements for non-interventional PASS and regulatory oversight.

4.1.1.23. Other study registration identification numbers and URLs as applicable (F1.1.1)

If the study is registered or published elsewhere, these identifiers can be included in this field. Any other links related to the study can be added here.

4.1.2. Study – Methodological aspects

4.1.2.1. Study topic (F8.1)

This field categorises the topic of the study, from the point of view of the medicinal product or disease/procedure studied:

- Disease/health condition
- Herbal medicinal product
- Human medicinal product
- Medical device
- Medical procedure
- Veterinary medicinal product

Where none of the values provided are suitable, further details can be provided in F8.1.1.

4.1.2.2. If 'other', further details on the study topic (F8.1.1)

Where the study topic is not covered by the available values, further details can be provided in this field.

*4.1.2.3. Study type * (F8.2)*

The type of study needs to be provided using this field. Based on the value selected here, the following fields are tailored to describe the specific study type provided.

- Clinical trial
- Non-interventional study
- Not applicable

Clinical trials in the scope of Directive 2001/20/EC should not be submitted to the Studies Catalogue.

Where 'Not applicable' is selected, further information should be entered using field F8.2.1

4.1.2.4. If 'Not applicable', further details on the study type (F8.2.1)

Where the study type is not covered by the available values, further details can be provided in this field.

4.1.2.4.1. Clinical trial regulatory scope (F8.2.2)

This field is applicable for studies where *'clinical trial'* was selected in field study type (F8.2).

From a regulatory standpoint, the following values are proposed (only one value can be selected):

- Clinical trial not part of marketing authorisation application or subject to marketing authorisation approval
- Post-authorisation interventional clinical trial
- Pre-authorisation clinical trial

Where none of these values are applicable, no value should be selected in this field.

4.1.2.4.2. Phase of the clinical trial (F8.4)

This field is applicable for studies where *'clinical trial'* was selected in field study type (F8.2).

From the point of view of regulatory lifecycle of a medicinal product, the following values are proposed:

- Human pharmacology (Phase I)
- Therapeutic confirmatory (Phase II)
- Therapeutic exploratory (Phase III)
- Therapeutic use (Phase IV)

Where none of these values are applicable, no value should be selected in this field.

4.1.2.4.3. Clinical trials randomisation (F8.3)

This field is applicable for studies where *'clinical trial'* was selected in field study type (F8.2).

The type of randomisation available for selection (only one value can be selected):

- Randomised clinical trial
- Non-randomised clinical trial

Where none of these values are applicable, no value should be selected in this field.

4.1.2.4.4. Clinical trial types (F8.3.1)

This field is applicable for studies where *'clinical trial'* was selected in field study type (F8.2).

A further classification is available as follows:

- Cluster randomised trial
- Large simple trial
- Low-intervention clinical trial
- Pragmatic clinical trial
- Single-arm trial

Where none of these values are applicable, no value should be selected in this field.

4.1.2.4.5. Non-interventional study design (F8.3.2)

This field is applicable for studies where *'non-interventional study'* was selected in field study type (F8.2).

The following choices are available (more than one value can be selected):

- Case-control

- Case-only
- Cluster design
- Cohort
- Cross-sectional
- Ecological
- Systematic review and meta-analysis

Where none of the above values are applicable, the value ‘Other’ should be selected. Further information can then be provided using field F8.3.2.1.

4.1.2.4.6. If ‘other’, specify design of non-interventional study (F8.3.2.1)

Where ‘Other’ is selected in field study design (F8.3.2), it is strongly recommended to use this field to provide further details on the design of the non-interventional study used. This information will be used to enrich the available look-up table and maintain structured data in the Catalogues.

4.1.2.4.7. Scope of the study (F8.6)

This field is applicable for studies where *‘non-interventional study’* was selected in field study type (F8.2).

The following values are proposed for the scope of the study. More than one value can be selected.

- Assessment of risk minimisation measure implementation or effectiveness
- Disease epidemiology
- Drug utilisation
- Effectiveness study (incl. comparative)
- Feasibility analysis
- Healthcare resource utilisation
- Hypothesis generation (including signal detection)
- Method development or testing
- Evaluation of patient-reported outcomes
- Safety study (incl. comparative)
- Scoping review (including literature review)
- Validation of study variables (exposure outcome covariate)

Where none of the above values are applicable, the value ‘Other’ should be selected. Further information can then be provided using field F8.6.1.

4.1.2.4.8. If ‘other’, further details on the scope of the study (F8.6.1)

Where ‘Other’ is selected in field scope of the study (F8.6), it is strongly recommended to use this field to provide further details on the scope of the study. This information will be used to enrich the available look-up table and maintain structured data in the Catalogues.

4.1.2.4.9. Data collection methods (F8.5)

This field is applicable for studies where *‘non-interventional study’* was selected in field study type (F8.2).

The following values are suggested to describe the data collection methods:

- Primary data collection
- Secondary use of data
- Combined primary data collection and secondary use of data
- No individual level data collected for the purpose of the study

Where none of these values are applicable, no value should be selected in this field.

4.1.2.5. Name of medicine (brand names of the medicines studies) (F17)

Where the study concerns a specific brand name or active substance, this should be provided here.

The look-up offers values synchronised with the available data on authorised medicinal products in EU, as available in Art 57 database. More than one value can be selected if needed.

If the value needed cannot be found in the look-up, it should be provided using field F17.1

4.1.2.6. Medicinal product information – other (F17.1)

Where the medicinal product information needed is not available in the look-ups provided (e.g.: brand name, active substance or ATC code) this field can be used to provide the required data.

4.1.2.7. Study drug International non-proprietary name (INN) or common name (F16)

Where the study concerns one or more specific medicinal products, the active substance can be provided using this field. The look-up offers values synchronised with the available data on substances as provided by the SMS service. More than one value can be selected if needed.

This field should be provided additionally to the brand name.

If the value needed cannot be found in the look-up, it should be provided using field F17.1

4.1.2.8. Anatomical Therapeutic Chemical (ATC) code (F15)

Where the study concerns one or more specific medicinal products, the ATC code can be provided using this field. The look-up offers values synchronised with the data provided by the RMS service, acting as a broker in delivering the WHO ATC code classification. More than one value can be selected if needed.

This field should be provided, if known, additionally to the brand name (F17) and INN/common name (F16).

If the value needed cannot be found in the look-up, it should be provided using field F17.1

4.1.2.9. Medical condition to be studied (F18)

Where the study concerns one or more specific medical condition, specifics can be provided using this field. The look-up offers values based on the data provided by the RMS service, acting as a broker in delivering the ICH MedDRA dictionary. More than one value can be selected if needed.

If the value needed cannot be found in the look-up, it should be provided using field F18.1.

4.1.2.10. Additional medical condition(s) (F18.1)

Where the medical condition value needed is not available in the look-ups provided, this field can be used to provide the required data.

4.1.2.11. Population studied: A short description of the study population (F2.5)

The description of study population should be provided using this field.

The field is limited to 10,000 characters.

4.1.2.12. Age groups (F2.5.1)

In characterising the age groups applicable to the study, the following categories should be used:

- in utero;
- paediatric population (< 18 years):
    - neonate:
        - preterm newborn infants (0 - 27 days);
        - term newborn infants (0 - 27 days);
    - infants and toddlers (28 days - 23 months);
    - children (2 to < 12 years);
    - adolescents (12 to < 18 years);
- adult and elderly population (>18 years):
    - adults (18 to < 65 years):
        - adults (18 to < 46 years);
        - adults (46 to < 65 years);
    - elderly (⩾ 65 years):
        - adults (65 to < 75 years);
        - adults (75 to < 85 years);
        - adults (85 years and over).
-

The look-up offers values based on the data provided by the RMS service, where a centralisation across EMA databases is performed, for the purpose of interoperability and ease of searchability. Therefore, other values are also available to select (i.e.: egg, embryonated egg, young animal, adult animal) – these values should not be used for the current use of the studies Catalogue. Future development might extend the scope of the Catalogues (e.g.: supporting veterinary medicine use cases) and make use of such values.

4.1.2.13. Special population of interest (F8.9)

The following categories are suggested to characterise special population of interest (more than one value can be selected):

- Frail population
- Hepatic impaired
- Immunocompromised
- Nursing women
- Pregnant women
- Renal impaired
- Women of childbearing potential not using contraception
- Women of childbearing potential using contraception

Where none of the above values are applicable, the value ‘Others’ should be selected. Further information can then be provided using field F8.9.1.

4.1.2.14. If ‘other’, specify which other population has been studied (F8.9.1)

Where the ‘special population’ (F8.9) is not covered by the available values, further details can be provided in this field.

4.1.2.15. Estimated number of subjects (F2.5.2)

An estimated number of subjects should be provided using this field.

4.1.2.16. Study design (brief summary of the study design) (F2.13)

A brief summary of the study design should be provided using this field. The text length is limited to 300 characters. Further information (e.g.: objective, setting) can be provided using the following fields.

4.1.2.17. Main study objective (short description of the study objective) (F8.10)

The main study objective should be provided using this field. The text length is limited to 10,000 characters.

4.1.2.18. Setting (F2.11)

The setting in terms of persons, place, time period and selection criteria, including a split by treatment arms/comparators or another relevant variable should be provided using this field. The text length is limited to 2000 characters.

4.1.2.19. Interventions (F2.6)

This field is applicable for studies where *'clinical trial'* was selected in field study type (F8.2) and should be used to describe the interventions used in the study. The text length is limited to 2000 characters.

4.1.2.20. Comparators (F2.7)

Where applicable, the comparators should be described using this field. The text length is limited to 2000 characters.

4.1.2.21. Outcomes (F2.8)

A brief summary of the outcomes, upon the study completion. The text length is limited to 2000 characters.

4.1.2.22. Data analysis plan (F2.12)

A brief summary of the analysis method (e.g. risk estimation, measures of risk, internal/external validity). The text length is limited to 2000 characters.

4.1.2.23. Summary results (F6.3)

A brief summary of the results of the study completion from the abstract should be provided here. The text length is limited to 2000 characters.

4.1.3. Study – Data management

4.1.3.1. Data sources (F3.4)

The data source(s) used in the study should be linked here. The linkage between data sources and the respective studies is an essential element in the functionality of the Catalogues.

4.1.3.2. Data sources, if not available in the list above (F3.4)

Where the data source(s) is (are) not available in the look-up provided, it (they) can be described here. The name of the data source should be accompanied by any available reference, wherever possible (e.g.: official website, publicly available contact points or details).

4.1.3.3. Data sources (types) (F8.7)

The data source classification can be provided using this field. More than one option can be selected:

- Administrative healthcare records (e.g. claims)
- Biobank
- Birth registry
- Cancer registry
- Clinical trial
- Congenital anomaly registry
- Death registry
- Disease registry
- Drug prescriptions
- Drug registry
- Electronic healthcare records (EHR)
- Expanded access program (compassionate use)
- Induced terminations registry
- Laboratory tests and analyses
- Mobile Health (mHealth)
- Non-interventional study
- Other
- Patient surveys
- Pharmacy dispensing records
- Population registry
- Pregnancy registry
- Published literature
- Social media
- Spontaneous reports of suspected adverse drug reactions
- Vaccination registry

4.1.3.4. If 'other', specify data sources type: (F8.7.1)

Where the data source category listed in field F8.7 does not cover the relevant data source, further information can be offered using this field.

4.1.3.5. Compliance with ENCePP Code of Conduct (F9.3.2)

This field can be used where the study is considered to have been performed in line with ENCePP Code of Conduct. Further information on this document can be found here.

4.1.3.6. CDM mapping: (F4.1)

This field should be marked as 'Yes' where the data source(s) in the study were converted (ETL-ed) to a CDM (common data model).

4.1.3.7. CDM name (F4.2)

Where field CDM mapping (F4.2) was marked as 'Yes', further information on the type of CDM used by the data source can be provided, using one of the available values:

Common Data Model	Description
BIFAP	Further reference here.

Common Data Model	Description
CDISC SDTM	Further reference here.
ConcepTION CDM	Further reference here.
CTcue Datamodel	Further reference here.
EUROCAT	Further reference here.
i2b2	Further reference here.
NorPreSS	A common data model developed for the Nordic Pregnancy Drug Safety Studies. Further reference here.
OMOP	The observational Medical Outcomes Partnership Common Data Model. Further reference here.
PCORnet	The National Patient-Centred Clinical Research Network Common Data Model. Further reference here.
PEDSnet	Further reference here.
Sentinel	Sentinel Common Data Model (SCDM) lead by the Sentinel Operations Center (SOC) residing within the Harvard Pilgrim Health Care institute. Further reference here.
Vaccine Safety Datalink (VSD) Data Dictionary	Further reference here.
TrineTX	Further reference here.
EUROMEDICAT	Further reference here.

If none of the above values are applicable, the value 'Other CDM' can be selected.

4.1.3.8. CDM mapping version or version date (F4.3)

Where field CDM mapping (F4.2) was marked as 'Yes', further information on the version of the CDM used can be used using this field (free text)

4.1.3.9. CDM name (other) (F4.2.1)

Where the CDM name is not available in the provided look-up in field F4.2, this field can be used to provide the relevant information.

4.1.3.10. Data quality specifications

Data quality specifications listed below aim to capture information related to data quality checks performed on the data source used for the submitted study (where applicable). For further information on data quality concepts described below, refer to EMA Data Quality framework published here.

*4.1.3.11. Check conformance * (F5.3)*

The checks on conformance for a data source aim to establish:

- Format: whether data are expressed in the same way throughout a dataset (e.g., a dataset mixing dates represented as DD-MM-YYYY and MM-DD-YYYY will not be suitable for an integrated analysis).
- Semantic: whether the same value mean the same thing throughout a dataset. For instance, whether "anuria" means a condition of total cessation of urine production or the measurement of the amount of urine, or whether the same notion of a measure is intended to have the same precision throughout a dataset.

*4.1.3.12. Check completeness * (F5.4)*

Completeness measures the amount of information available with respect to the total information that could be available given the capture process and data format. Data unavailable in the dataset (either due to systematic reasons such as information available in the data source but not included in the data model, or specific entries that are unavailable for a given field) are called "missing". For example, the percentage of non-missing values for a required field (e.g., sex) in a dataset would be a measure for completeness.

*4.1.3.13. Check stability * (F5.5)*

A check whether the same entities are identified in the same way throughout a dataset and the conventions used haven't changed in time.

*4.1.3.14. Check logical consistency * (F5.6)*

A check of logical consistency, aiming to characterise the reliability of a data source. This concept, described in EMA Data Quality Framework as plausibility, refers to checks such as:

- Data values and distributions agree with internal measurements or local knowledge (e.g.: height and weight are a positive value);
- Logical constraints between values agree with common knowledge;
- Data values and distributions agree with trusted reference standards.

Further examples of the concept can be found here.

*4.1.3.15. Data characterisation conducted * (F5.1)*

Where a data characterisation was conducted or a quality check process was completed, this field should be marked as 'Yes'.

4.1.3.15.1. Data characterisation moment (F5.2)

Where field 'Data characterisation' (F5.1) is marked as 'Yes' further information on the stages of the study where data characterisation steps or quality checks were implemented can be provided using the following values (more than one value can be selected):

- after data extraction
- after extract-transform-load to a common data model
- after creation of study variables

4.1.3.15.2. Data characterisation details (F5.7)

Where field ‘Data characterisation’ (F5.1) is marked as ‘Yes’ further information on the summary description of the data characterisation or quality check process can be provided here. The text field length is limited to 1000 characters.

4.1.3.15.3. Data characterisation results (F5.8)

Where field ‘Data characterisation’ (F5.1) is marked as ‘Yes’ further information on the results of the data characterisation or quality checks can be provided using this field (e.g.: OMOP/OHDSI data quality dashboard or the Sentinel Common Data Model level 1-4 checks). The information can be uploaded as a pdf file (which will subsequently be made public) or as a link reference to information published elsewhere.

Pdf file size is limited to 20Mb.

4.1.3.16. Procedure of data extraction (F6.1)

The procedure of data extraction can be provided using this field.

The information can be uploaded as a pdf file (which will subsequently be made public) or as a link reference to information published elsewhere.

Pdf file size is limited to 20Mb.

4.1.3.17. Procedure of result generation (F6.2)

The procedure of result generation can be provided using this field.

The information can be uploaded as a pdf file (which will subsequently be made public) or as a link reference to information published elsewhere.

Pdf file size is limited to 20Mb.

4.1.4. Study – Resources

*4.1.4.1. Protocol link * (F2.1)*

A link to the latest version of the protocol, if already published elsewhere.

4.1.4.2. Protocol document (F2.3)

The protocol document can be uploaded in this field. While the field is not marked as mandatory, the upload of this document is strongly encouraged.

The study protocol should be provided before the start of data collection. Where prior publication of the protocol could threaten the validity of the study or the protection of intellectual rights, a study protocol with redactions may be entered into the register prior to the start of data collection. Further information about the requirements for the registration of PASS is available in the guideline on Good Pharmacovigilance Practices (GVP) module VIII, chapter VIII.B.4.

There is no limit to the number of versions of the study protocol that can be uploaded in the system. No changes can be made to the "initial" document throughout the history of the study record once it has been uploaded and submitted. Subsequent versions of the document can be overwritten as often as necessary with a newer version, but only the very latest version will be visible.

The pharmacovigilance legislation requires the EMA to publish in a publicly available register (the HMA/EMA studies Catalogue) the protocols and abstracts of results of PASS imposed as an obligation by a competent authority in accordance with Articles 10 or 10a of Regulation (EC) No 726/2004 or with Articles 21a or 22a of Directive 2001/83/EC. It also specifies that the final report of such imposed PASS must provide the date of registration in this register.

4.1.4.3. Study results (F6.4)

The study results should be uploaded in pdf format using this field. The uploaded document will be made public, to support transparency.

4.1.4.4. Study report (F7.2)

The study report should be uploaded in pdf format using this field. The uploaded document will be made public, to support transparency.

4.1.4.5. Study, other information (list of URLs to other relevant resources describing the study and/or files as applicable) (F7.3)

Any other files and/or links can be added using this field. The field can be used to submit more files and/or links, as needed.

4.1.4.6. Study publications (F7.1)

Links to peer-reviewed papers reporting the study can be provided using this field.

4.1.4.7. ENCePP Seal: requesting ENCePP seal for the study (F9)

Where an ENCePP Seal is requested, this field should be marked as 'Yes'

Further documentation on the process and requirements surrounding the assignment of an ENCePP seal can be found here.

4.1.4.8. ENCePP Seal granted (F9)

This data field is used by EMA during the validation process to indicate studies where the ENCePP Seal has been granted.

4.1.4.9. ENCePP Seal relevant documents

The following documents should be uploaded, in pdf format, where an ENCePP Seal is being requested (field F9 marked as 'Yes'):

- Conflicts of interest of investigators (F9.1)
- Composition of steering group and observers (F9.2)
- Signed code of conduct (F9.3)
- Signed code of conduct checklist (F9.3.1)
- Signed checklist for study protocols (F9.4)

Further documentation on the process and requirements surrounding the assignment of an ENCePP seal can be found here.

5. Registering a Study

Any researcher, sponsor or regulator may submit a(n) (observational) study entry in the HMA-EMA Catalogue of real-world data studies, regardless of the country(ies) in which the studies are conducted.

The questionnaire consists of 23 questions and are divided into three categories:

1. Administrative details,
2. Data elements collected, and
3. Data flows and management.

All mandatory fields, marked with a red asterisk (*), will need to be completed in order to move to the next step. A sample questionnaire for offline review can be found on the support page. A draft can only be saved once all the mandatory fields have been filled in. We strongly encourage to fill in as many fields as possible, and to be as descriptive and detailed as possible in the description fields. Refer to section 4. Catalogue of real-world data studies: Description of the data fields and definitions for further explanations and definitions on the metadata fields.

To enter a study in the RWD Catalogue, follow the link: Catalogue of RWD studies | HMA-EMA Catalogues of real-world data sources and studies (europa.eu). Ensure that you are logged in, using your EU Login account credentials. If you do not have an EU Login account, visit the support page on how to create one. Any designated individual (e.g., staff member of the research centre, sponsor or pharmaceutical company acting as study funder) may register the study with the agreement of the primary lead investigator. The decision of who enters the study in the Catalogue lies with the primary lead investigator who is ultimately responsible for the published information. The individual registering a study must have an EU Login account. Multiple users may collaborate on a single study record, as needed. Please see section 8.6. 8.6. Adding a co-author on how to submit a collaboration request.

Registration of non-interventional post-authorisation safety studies (PASS) conducted pursuant to an obligation imposed by a competent authority, in accordance with Articles 10 or 10a of Regulation (EC) No 726/2004 or with Articles 21a or 22a of Directive 2001/83/EC, is mandatory in line with the provision of GVP Module VIII on PASS. According to Article 26(1)(h) of Regulation (EC) No 726/2004, protocols and public abstracts of results of non-interventional post-authorisation safety studies (PASS) relating to authorised medicinal products imposed as an obligation of marketing authorisation by a competent authority in accordance with Article 10 or 10a of Regulation (EC) No 726/2004 or with Articles 21a or 22a of Directive 2001/83/EC shall be made public. MAH(s) should follow the guidance and timelines for the publication of study documents as outlined in the GVP Module VIII.

In order to support transparency on all non-interventional post-authorisation safety studies (PASS) and to facilitate exchange of pharmacovigilance information between the Agency, regulatory authorities and marketing authorisation holders, the marketing authorisation holders should also enter in the HMA-EMA Catalogue of RWD studies all non-interventional PASS conducted voluntarily in the EU or included in the risk management plan agreed in the EU. Further information about the requirements for the registration of PASS is available in the guideline on Good Pharmacovigilance Practices (GVP) module VIII, chapter VIII.B.

EMA also strongly encourages MAHs, researchers and regulatory bodies to upload the final study results of non-interventional (PAS) studies in the Catalogue of RWD studies to support transparency and to facilitate exchange of pharmacovigilance information between the EMA, NCAs and MAHs.

Information on post-authorisation efficacy studies (PAES) that are not clinical trials (i.e., outside the scope of Directive 2001/20/EC) should also be entered in the Catalogue of RWD studies to

support transparency on post-authorisation efficacy studies (PAES), whether they are initiated, managed or financed by a marketing authorisation holder voluntarily or pursuant to an obligation.

The registration and publication of study documents should be performed in the Catalogue of RWD studies. Additionally, the registration of studies that wish to qualify for the ENCePP Seal is mandatory.

In addition to registering imposed studies, the Catalogue of RWD studies is also open to all RWD studies, that classify as interventional studies (e.g., large sample trial, pragmatic clinical trials etc.). This also includes all studies managed by marketing authorisation holders, research centres and organisations. Registration of studies is strongly encouraged and can be done at any stage (i.e., from planned to already finalised studies).

Please note, the Editor will receive email notifications from the following domain: fpfis.tech.ec.europa.eu. If you have not received a notification, please check your spam inbox or contact us.

5.1. Validation and publication process

EMA is responsible for reviewing and validating the content that is submitted for publication in the Catalogue of RWD studies. The EMA Validator uses the following criteria to assess the submission:

- Non-Duplication: The study record should not duplicate another existing record in the Catalogue.
- Exclusion of Clinical Trials outside the scope of the Catalogue: In case of a clinical trial, the EMA Validator needs to verify that the study falls within the scope of the Catalogue. Clinical trials that are outside the scope of the Catalogue will be excluded.
- Relevance: the study should include a medicinal product/device or, a class or a specific disease or disease area.
- Required documentation for PASS imposed studies: protocols and abstract of results (as applicable for the study timeline) should be published.

If the submission does not meet all the validation criteria described, the EMA Validator will return the submission using the Catalogues' data management notification system, adding a message in the Revision Log containing the justification for the rejection/return. Subsequently, the Editor may further refine the submission with additional information (as required) and resubmit. This process is repeated until all the criteria have been met.

Once all the criteria have been met, the EMA Validator will verify that the study submission is *complete*, meaning that all relevant fields (e.g. sources of funding, study description, age groups, main study objective(s), scope of the study, data analysis plan, name of medicine (if applicable), medical condition to be studied (if applicable), estimated number of subjects) are filled in and the information is consistent. In addition to the mandatory fields, the applicability of each data element depends on the type of study validated. The validation is performed against the submitted attachment files (e.g.: study protocol, study results where applicable) and/or publicly available information (e.g.: publications). An assessment of the *accuracy* of information and internal validity is also performed, using similar references as above. These checks aim to ensure high quality of the structured information provided with the study data. The *timeliness* of the submissions of the imposed PASS are also assessed in accordance with the guidance set out by GVP Module VIII. Additionally, the uploaded documentation will also undergo validation. Documents containing internal comments or pending tracked changes will be returned for resubmission.

Regarding the format and content of the study protocol, abstract and final study report the Implementing Regulation 520/2012 provides in Annex III a clear structure which MAHs should follow when submitting protocol, abstract and final study reports to regulatory authorities for PASS imposed as an obligation. To

this end the Agency has published guidance for the submission of PASS protocols on the EMA website: https://www.ema.europa.eu/documents/other/guidance-format-content-protocol-non-interventional-post-authorisation-safety-studies_en.pdf. Similar guidance is provided for abstracts and final study reports: https://www.ema.europa.eu/documents/regulatory-procedural-guideline/guidance-format-content-final-study-report-non-interventional-post-authorisation-safety-studies_en.pdf. To ensure the same level of transparency and, scientific and quality standards for all PASS, MAHs should also follow the format and content requirements when making information about PASS available in the Catalogue of RWD studies. However, this is a recommendation and not a legal requirement as applicable to PASS conducted pursuant to an obligation.

If the submission is considered incomplete or inaccurate, the EMA Validator may request the user to make the necessary changes, or to provide further clarifications. The EMA Validator may also perform minor corrections, where necessary. Minor corrections, in this instance, refer to minor adjustments made by the EMA Validator to maintain data consistency, primarily from a data management perspective. These changes are meant to streamline the submission of information without altering its content and meaning. For example, misplacing of information in the incorrect field (e.g., study scope information provided in the study results field) would qualify as a minor error and can be corrected by the EMA Validator. Any other inaccuracies will be considered major and will be returned to the Editory for correction.

If the study entry passes the EMA assessment, the EMA Validator approves the study which triggers the automatic publication of the study entry in the Catalogue of real-world data studies and the entry is now available for public viewing.

5.2. Maintenance of a study

It is important that the metadata information published in the Catalogue of real-world data studies is kept up-to-date. Either the primary lead investigator or an appointed study administrator is expected to keep the information of their respective study up to date on a regular basis. It is strongly encouraged to update the record at each study milestone (i.e. date when funding contract was signed, study start date, data analysis start date, date of interim report if expected, and date of final study report). At a minimum, the user will receive notifications via email to remind them of the following study milestones, 30 days after the indicated planned date: start of data collection, and planned date of final study report. Of particular importance, and to support transparency and the exchange of information, the maintenance of non-interventional PASS conducted pursuant to an obligation by an EU competent Authority should follow the milestones and guidance described in GVP Module VIII. Updated study protocols in case of substantial amendments, progress reports and the final study report should also be entered in the Catalogue of RWD studies as soon as possible, and preferably within two weeks after their finalisation.

To update the study record, the user will need to login in order to access their dashboard which contains the list of entries associated to their account. The user can edit the entry at any time. Once the information in the entry is updated, the entry will need to be submitted for validation once more before the update can be published. The EMA Validator will carefully review and assess the updated information according to the criteria and steps outlined in section 5.1. Validation and publication process.

If user wishes for the study to be unpublished/removed from the Catalogue of RWD studies, the user may send a request to the EMA validator to unpublish the study record, providing sufficient justification for this request. Please note that while there is no provision in GVP Module VIII or in the legislation that determines if or when study records may be unpublished/removed from the RWD Catalogues, EMA would favour to keep the study record published in the RWD Catalogues in the interest of transparency. By adding a note in the field ‘Brief description of the study’, you may indicate if a study was terminated prematurely or cancelled, including relevant explanatory information. If such a request for removal is

submitted, the EMA Validator will carefully consider the request and justification. The timelines to process this request may take up to two weeks or up to one month in exceptional circumstances.

Once a study record is updated and submitted, the EMA validator will review the updated record. If satisfactory, the EMA Validator will approve the resubmission of the study entry, thereby triggering the automatic publication of the study entry and its updates in the Catalogue. If the study is finalised, the system will automatically change the study status as “finalised” once the field ‘Actual date of final study report’ is filled in by the Editor.

6. Registering an Institution

The HMA-EMA Catalogues also contains a section which covers institutions. These are organisations in the healthcare sector that can either hold data sources or perform research based on real-world data related to these sectors. Any institution can register in this Catalogue.

The questionnaire consists of four questions. Mandatory fields are marked with a red asterisk (*). A sample questionnaire for offline review can be found on the support page. A draft can only be saved once all the mandatory fields have been filled in. We strongly encourage to fill in as many fields as possible, and to be as descriptive and detailed as possible in the description field.

To enter an institution in the RWD Catalogue, follow the link: Institutions | HMA-EMA Catalogues of real-world data sources and studies (europa.eu). Ensure that you are logged in, using your EU Login account credentials. If you do not have an EU Login account, visit the support page on how to create one.

An institution ID will be automatically generated. Provide the institution’s full name and acronym in the first field. Select the country/ies where the institution head office or coordinating centre is located. From the drop-down list, select the sector where the institution operates (i.e., not-for-profit, regulatory authority, pharmaceutical company etc.). From the drop-down list, select the institution’s role in connection to the HMA-EMA Catalogues of real-world data sources and studies (i.e., data holder, data providers, researcher, other). Provide the link to the institution website and add a description of max 2000 characters about the institution.

Select whether your institution collects data directly from individual patients/respondents, and/or whether there is any interest in carrying out research funded by pharmaceutical companies. Also select whether your institution is interested in becoming an ENCePP Partner. More information on becoming an ENCePP Partner is available on the ENCePP website.

Lastly, provide a full name and functional email address which will serve as a main point of contact. These details will be made public once the entry is validated and published.

The validation process of an institution is similar to the validation process of a data source (see section 3.1. 3.1. Validation and publication process). Once the institution entry passes the EMA assessment, the EMA Validator approves the institution which triggers the automatic publication of the institution entry in the Catalogue of institutions and the entry is now available for public viewing.

Should any updates be required over time, the entry will be available in ‘My dashboard’ for editing and resubmission.

Please note, the Editor will receive email notifications from the following domain: fpfis.tech.ec.europa.eu. If you have not received a notification, please check your spam inbox or contact us.

7. Registering a Network

The HMA-EMA Catalogue also contains a section which covers networks in the health or care sector. Networks are virtual structures defined by a formal agreement between individuals, organisations and/or structures sharing and collaborating towards the same objectives and quality standards. Any network can register in this Catalogue.

The questionnaire consists of five questions. Mandatory fields are marked with a red asterisk. A sample questionnaire for offline review can be found on the support page. A draft can only be saved once all the mandatory fields have been filled in. We strongly encourage to fill in as many fields as possible, and to be as descriptive and detailed as possible in the description field.

To enter a network in the RWD Catalogue, follow the link: Create Network | HMA-EMA Catalogues of real-world data sources and studies (europa.eu). Ensure that you are logged in, using your EU Login account. If you do not have an EU Login account, visit the support page on how to create one.

A network ID will be automatically generated. Provide the network's full name and acronym in the first field. Select the country/ies where the network is located. Provide the link to the network's website and add a description of max 2000 characters about the network. From the drop-down list, select the primary therapeutic/disease areas of the network. Also select from a drop-down list what funding sources the network has (e.g., EMA, NCAs, non-for-profit organisations, public funding, pharmaceutical company etc.).

Select whether your network is interested in becoming an ENCePP Partner. More information on becoming an ENCePP Partner is available on the ENCePP website.

Provide a full name and functional email address which will serve as a main point of contact. These details will be made public once the entry is validated and published.

Lastly, select from the drop-down list which institution(s) are part of this network. The drop-down list contains all institutions indexed in the RWD Catalogues. If the institution you are looking for is not in the RWD Catalogues, enter the name in the free text field below.

The validation process of a network is similar to the validation process of a data source (see section 3.1. 3.1. Validation and publication process). Once the network entry passes the EMA assessment, the EMA Validator approves the network which triggers the automatic publication of the network entry in the Catalogue of networks and the entry is now available for public viewing.

Should any updates be required over time, the entry will be available in 'My dashboard' for editing and resubmission. The information published in the RWD Catalogues needs to be kept up-to-date by the record owner. It is not the responsibility of the EMA.

Please note, the Editor will receive email notifications from the following domain: fpfis.tech.ec.europa.eu. If you have not received a notification, please check your spam inbox or contact us.

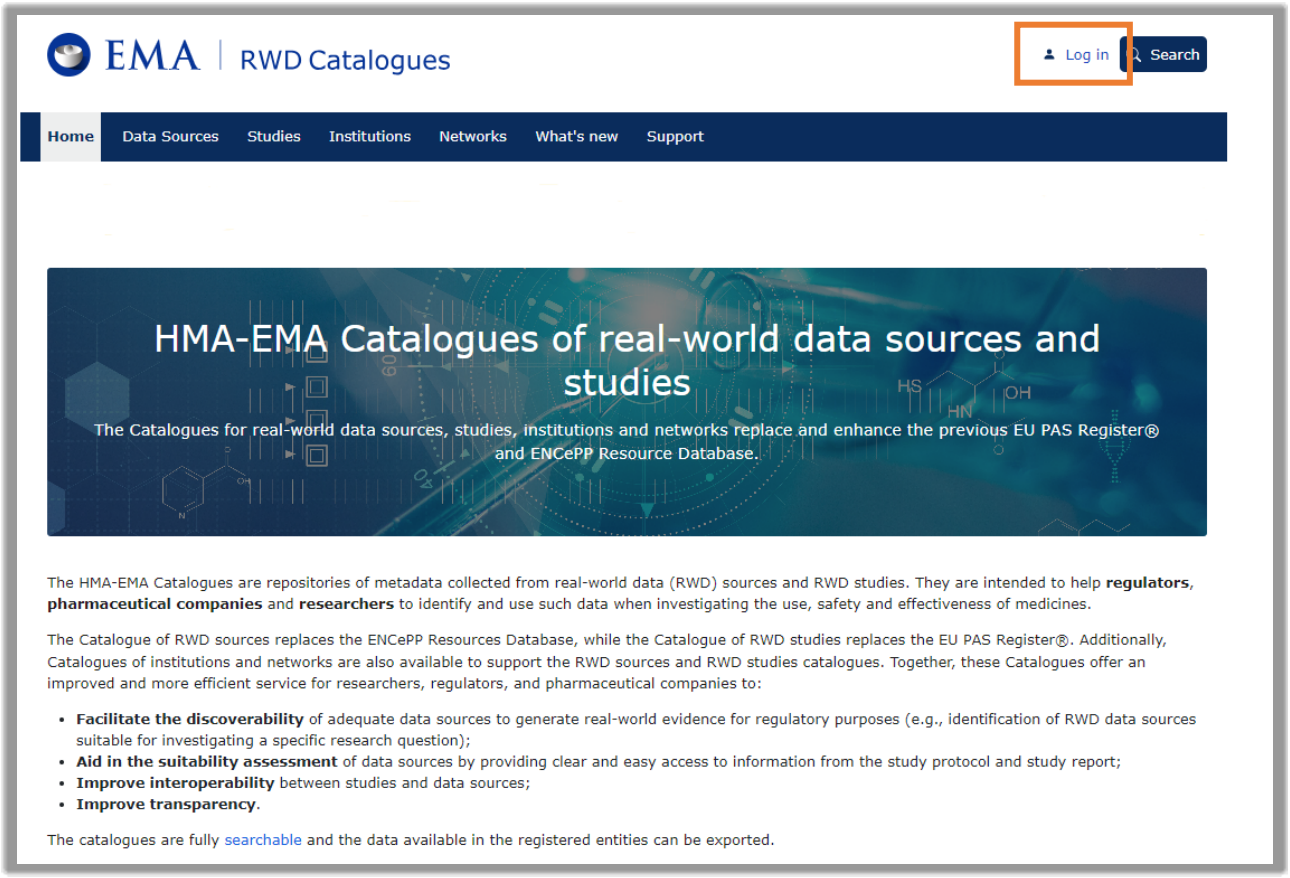
8. User guidance on how to navigate the RWD Catalogues website

8.1. Home page and log in

You can access the homepage of the RWD Catalogues website via the following link: https://catalogues.ema.europa.eu/

To log in, click on ‘Log in’ in the top right-hand corner of the webpage. Log in using your EU Login account credentials and follow the steps through the multi-factor authentication process. If you do not already have an EU Login account, you can create one here: EU LOGIN (europa.eu)

You do not need to log in to view and/or download the published entities in the RWD Catalogues. To view published entities, click on the ‘Search’ button in the top-right hand corner.

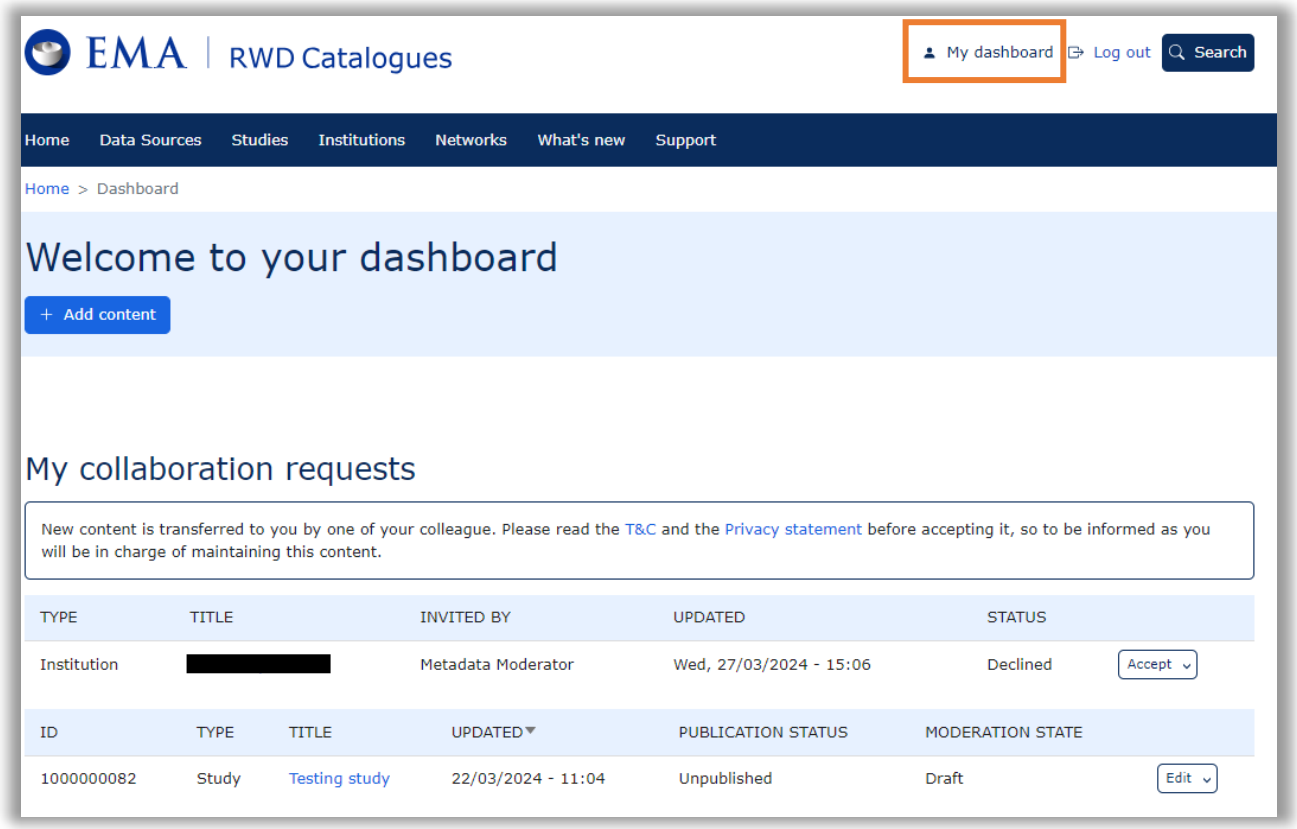


8.2. My dashboard

Once you have logged in to the RWD Catalogues, you will be able to access your dashboard.

To access your dashboard, click on ‘My dashboard’ in the top right-hand corner.

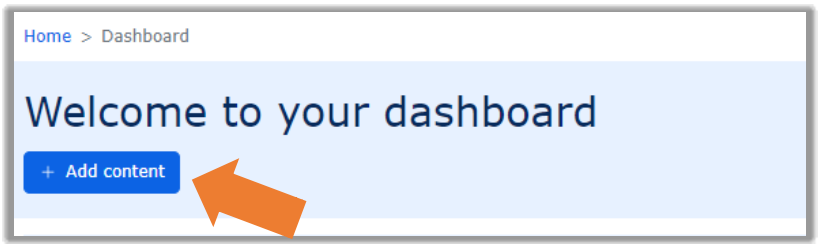
Your dashboard will contain the list of records that you own, as well as the list of records for which you have been requested to collaborate.



8.3. Add a record

To add a record, click on the '+ Add content' button in your dashboard.

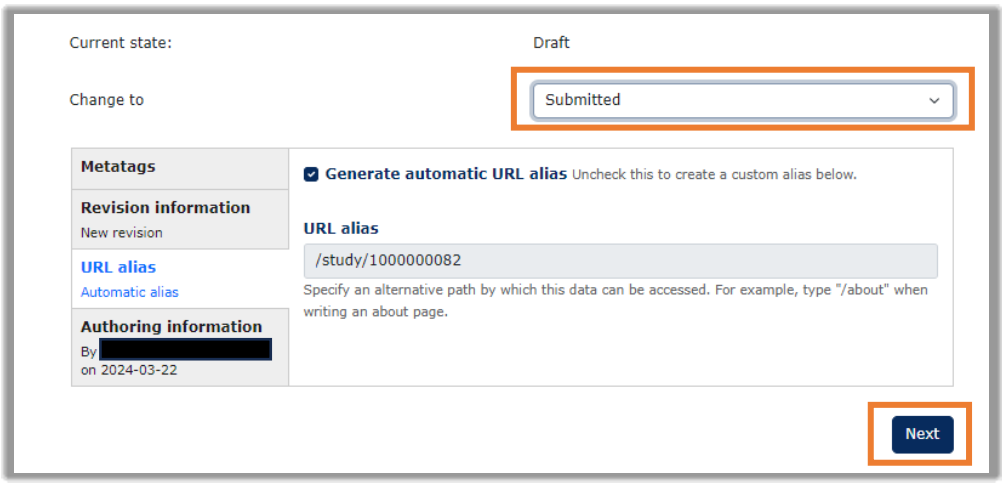
Select which type of entity you would like to enter in the RWD Catalogues: Data source, Institution, Network, or Study.



8.4. Submit a record for approval and publication

To submit a record for approval and publication, ensure that all mandatory fields are filled in. Mandatory fields are marked with a red asterisk (*). To submit, scroll down to the bottom of the record page (in 'Edit' mode) and change the record state from 'Draft' to 'Submitted'. Click on the 'Next' button. You will have to click 'Next' in all steps of the record (3 steps for a Study record, 4 steps for a Data source record and 1 step for Network and Institution records).

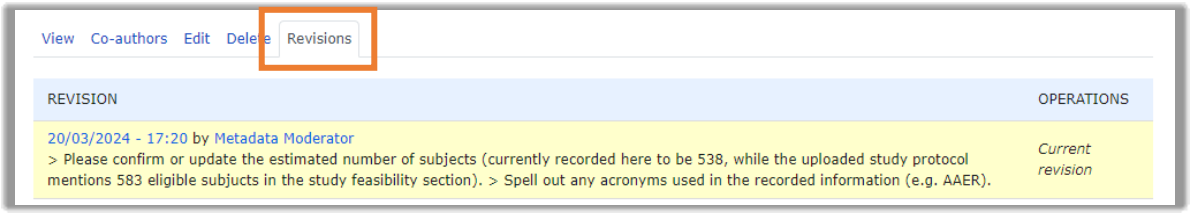
Please note that you will not be able to see the changes made to the record immediately in the 'View' mode. Once your record is submitted, EMA will take up to 10 working days to review your record. If your record requires corrections or more information, EMA will return the record and provide a comment in the Revision Log with the changes required. If the record is approved, it will automatically be published in the RWD Catalogues. EMA will notify you of the outcome of the review via email.



8.5. Record is returned

If your record is returned after submitting it for publication, you will receive an email indicating that your record is returned and needs further clarification/information.

In the 'Revisions' tab of your record, you will be able to see the comments from the EMA Validator (aka Metadata Moderator) regarding which aspects of the record need to be amended for it to be approved and published. Please make the necessary amendments and submit it once more for approval.



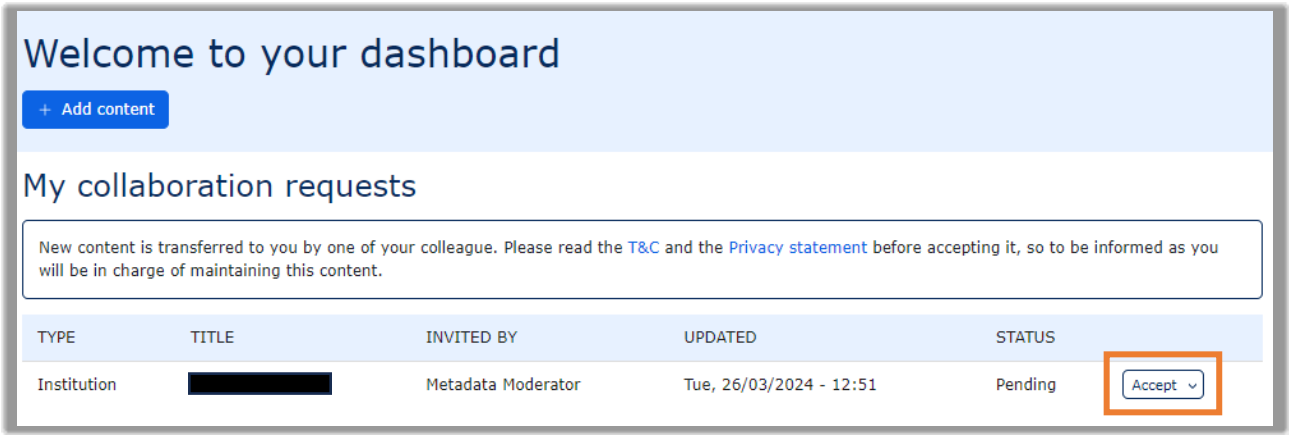
8.6. Adding a co-author

The RWD Catalogues allows for multiple users to collaborate on a single record. This can be done by adding a co-author via the 'Co-authors' tab in any record.

Enter the email address of the co-author you wish to add and slick 'submit'. Please ensure that the co-author you wish to add has an EU Login account and has logged in at least once in the RWD Catalogues website (to activate their account).

8.7. Accepting a collaboration request

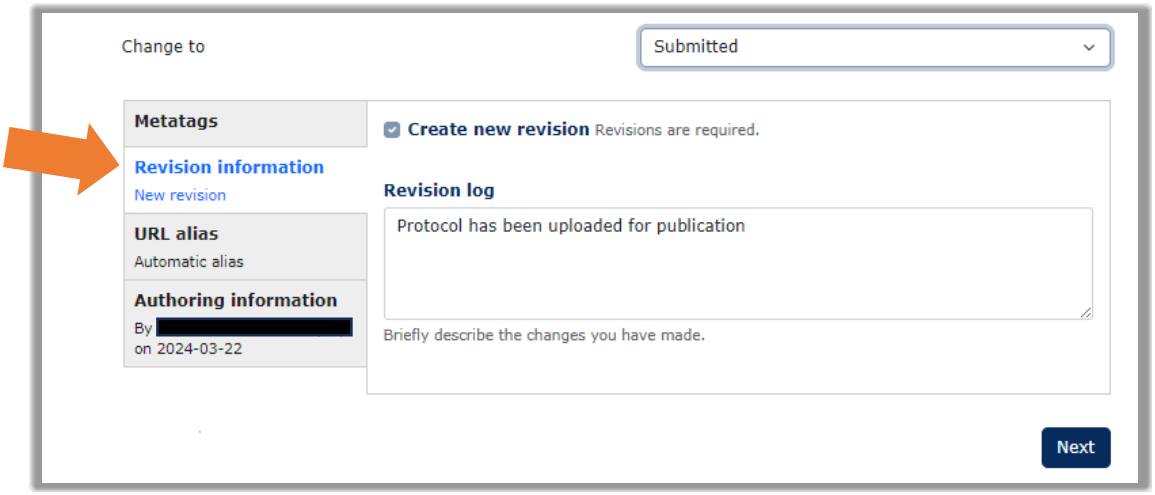
To view a collaboration request to co-author a record, you must log in to the RWD Catalogues and go to 'My dashboard'. In your dashboard, you will see a section 'My collaboration requests' where all requests to co-author a record will be listed. To accept a request, click on the 'Accept' button. Once accepted, you will be able to edit the record by clicking on the title hyperlink, and then selecting the 'Edit' tab.



8.8. Using the revision log

To facilitate EMA's review of a record that is already published and was updated, EMA requests that the user indicates which changes/updates were made in the 'Revision log' box when submitting.

For example, you may enter in the 'Revision log' that the protocol file has now been uploaded to the study record. EMA will then focus the review on the newly uploaded protocol.



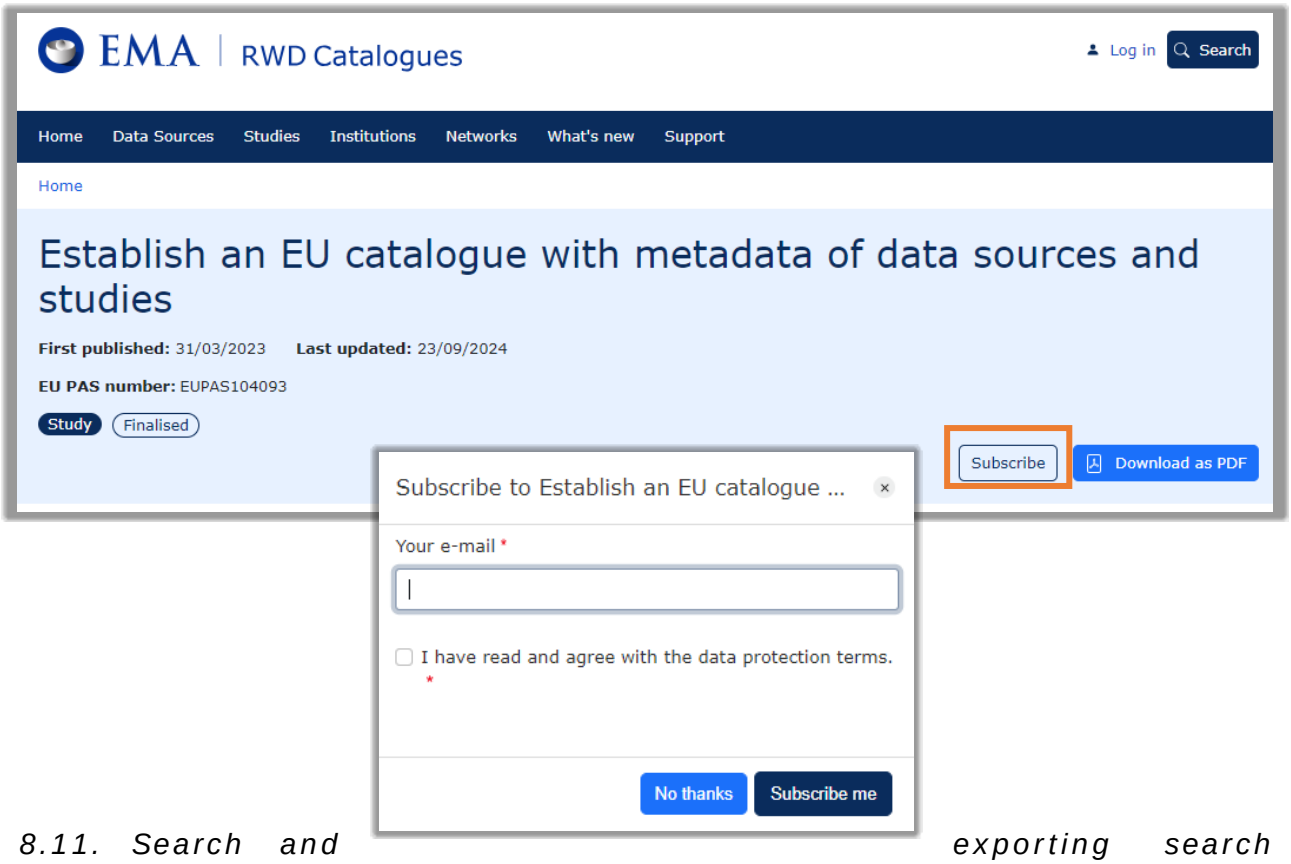
8.9. Downloading records

Any record (data source, study, institution, and network) can be downloaded as PDF from the RWD Catalogues website. Select the record you wish to download and click on the 'Download as PDF' button.

8.10. Subscribing to study records

Anyone can subscribe to a study record. This functionality is available to people who wish to stay updated on the status of a study. Once subscribed to a study record, the person will receive an email every time the study record is updated and published.

To subscribe, select a study record and click on the ‘Subscribe’ button. Fill in your email address in the pop-up box and tick to agree with the data protection terms.



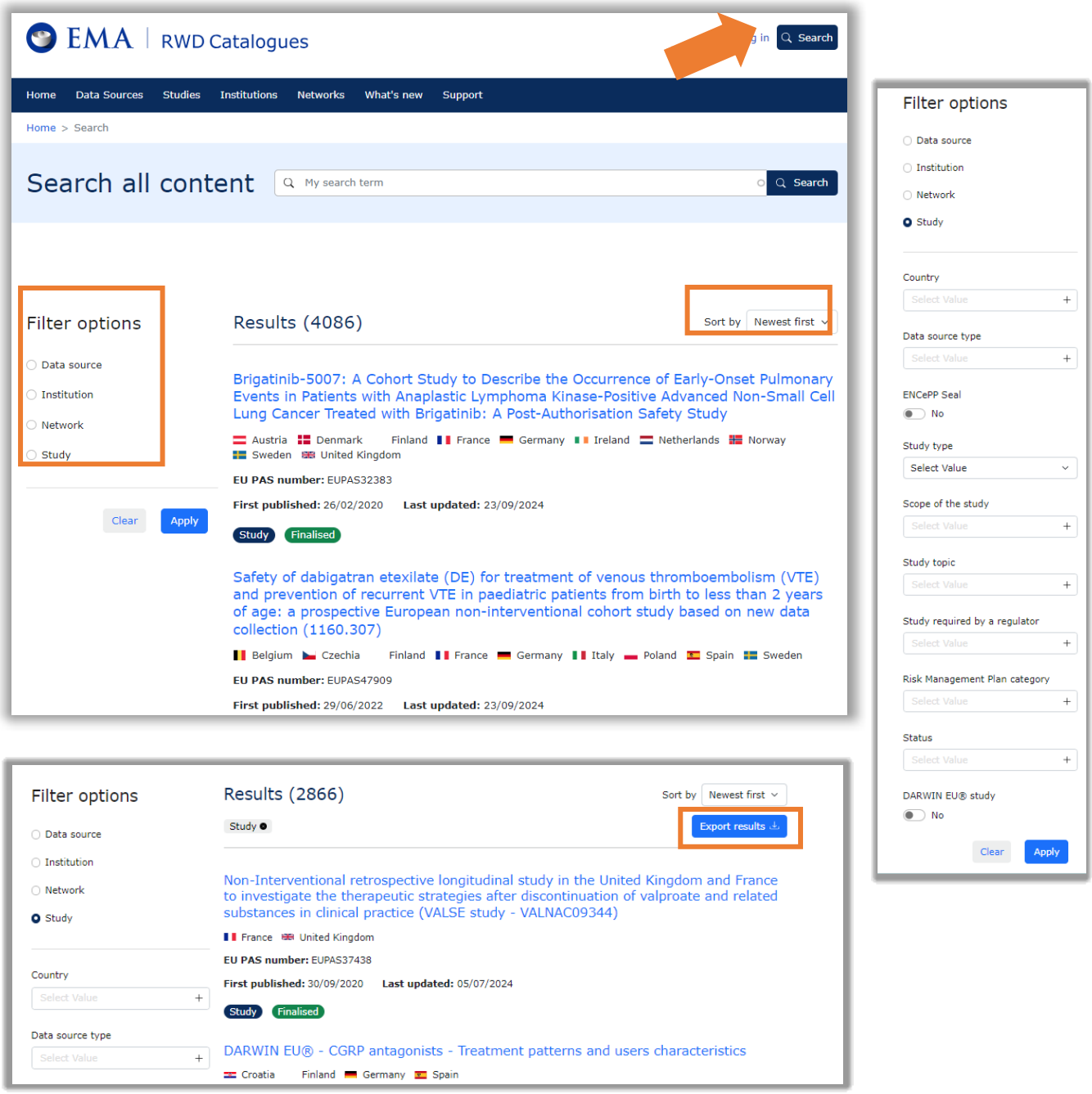
8.11. Search and exporting search results

Anyone can search the RWD Catalogues, by clicking on ‘Search’ in the top right-hand corner. You do not need to be logged in to search or export the search results.

The search results can be sorted by newest or oldest first, or in alphabetical order.

Filtering options are available for each type of entry on the left-hand side. Once a record type has been selected, further filter options will become available relevant to that record type.

Search results can be exported by clicking on the ‘Export results’ button. An Excel file will automatically start downloading.



9. Additional documentation and useful links

Homepage | HMA-EMA Catalogues of real-world data sources and studies (europa.eu)

Big data | European Medicines Agency (europa.eu)

ENCePP (europa.eu)