31 May 2022
EMA/563896/2022

List of metadata for Real World Data catalogues

Reviewed by European Network Data Board	28 April 2022
Adoption by Big Data Steering Committee	03 June 2022
Sent for information to Heads of Medicines Agencies	07 June 2022
Sent for information to EMA Management Board	07 June 2022
Sent for information to European Commission	07 June 2022

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 An agency of the European Union

Table of Contents

1. Introduction

In line with the BDSG workplan and the EMRN Strategy to 2025 action on data discoverability, and through consultation with its stakeholders, the EMA produced a list of metadata for describing real-world data (RWD) sources and studies.

The chosen metadata will be included in a catalogue of data sources containing information about existing real-world databases (to replace the current ENCePP catalogue) and information about the studies performed on the data sources (to replace and enhance the current EU PAS Register).

Setting up a catalogue of data sources and enhancing the catalogue of studies aims to improve transparency with regard to observational studies, discoverability of studies and data sources and contribute to increasing the ability to judge the evidentiary value of observational studies and real-world data sources when used to investigate the use, safety and effectiveness of medicinal products.

2. List of metadata

2.1. Data source metadata

I. Data source – Administrative details

1) Name of the data source (as used in European projects) *<free text>* (C1.2)
2) Data source acronym, if applicable *<free text>* (C1.3)
3) Data custodian *<free text | lookup of institutions>* (C4.1)
4) Data source contact name *<free text>* (M1.3)
5) Data source contact email *<free text> (*M1.6*)*
6) Data source countries (where data originates) <*select multiple:* ISO 3166-1 country codes> (C1.5)
7) Data source language(s) <*select multiple:* ISO 639 codes> (C6.2)
8) Data source regions (geographical regions that the data source covers) <*select multiple*: ISO 3166-2 country codes> (C1.5.1)
9) Date when the data source was first established *<date>* (C4.5)
10) Data source time span: First collection *<date>* (C1.12), Last collection *<date>* (C1.13)
11) Data source website (where applicable, a dedicated website for the data source) *<free text| weblink>* (C11.1)
12) Data source publications (A list of peer-reviewed papers or documents describing the data source (validation, data elements, representativity) or its use for pharmacoepidemiologic research): *<free text| weblink>* (C11.2)
13) Data source qualification: the data source has successfully undergone a formal qualification process (e.g., from the EMA, or ISO or other certifications) *Yes | No* (C3.1)

    If yes, description of the qualification *<free text> (C3.1.1)*
14) Main financial support of the data source in the last three years: <*select multiple:* Funding by own institution| national, regional, or municipal public funding| European public funding| funding from industry or contract research| funding from public-private partnership| funds from patients organisations, charity and foundations| other> (C4.6)
15) Data source type <*select multiple:* Hospital discharge records| primary care medical records| pharmacy dispensation records| birth registry| induced terminations registry| congenital anomaly registry| population registry| registration with healthcare system| outpatient visit records| emergency care discharge records| exemptions from co-payment| diagnostic tests or procedures reimbursement| death registry| cancer registry| other disease registry| vaccination registry| drug registry| biobank| administrative claims | spontaneous reporting of adverse drug reactions | electronic health records | specialist care records | other > (C5.1)

    If 'other', data source type: *<free text>* (C5.1.1)
16) Care setting for data source <*select multiple:* primary care – GP, community pharmacist level | primary care – specialist level (e.g. paediatricians) | secondary care – specialist level (ambulatory or hospital outpatient care) | hospital inpatient care | other> (C1.14)

II. Data source – Data elements collected

17) The data source contains information on the following (tick as applicable):

☐ Specific diseases (C1.10)
☐ Hospital admission discharge (C6.10)
☐ ICU admission (C6.10.1)
☐ Cause of death (C6.11)
☐ Rare diseases (C6.12)
☐ Prescriptions and/or dispensing (C6.13)
☐ ATMP (C6.16)
☐ Contraception (C6.17)
☐ Indication for use (C6.18)
☐ Administration of vaccines (C6.19)
☐ Administration of other injectables (C6.19.1)
☐ Medical devices (C6.20)
☐ Procedures (C6.21)
☐ Clinical measurements (C6.23)
☐ Healthcare provider (C6.24.1)
☐ Genetic data (C6.25)
☐ Biomarker data (C6.26)
☐ Patient-generated data (C6.27)
☐ Units of healthcare utilisation (C6.29)
☐ Unique identifiers for persons (C6.4)
☐ Diagnostic codes (C6.9)
☐ Pregnancy and neonates (C1.9)

18) For data sources collecting specific disease information – disease information collected *<free text>* (C1.10.1)

19) Population age groups (tick as applicable): (C1.8)

☐ newborn infants (0 to 27 days),
☐ infants and toddlers (28 days to 23 months),
☐ children (2 to 11 years),
☐ adolescents (12 to 17 years),
☐ adults (18 to 45 years),
☐ adults (46 to 64 years),
☐ adults (65 to 75 years),
☐ adults (76 to 85 years),
☐ adults (86 years and over),
☐ all ages

20) Family linkage available in the data source: <*select multiple:* None | Household | mother-child | father-child | sibling> (C6.6)

21) If not available, can familial linkage be created on an ad-hoc basis: <*select one:* Yes | No > (C6.6.1)

22) Sociodemographic information collected: <*select multiple:* None | age | gender | ethnicity | country of origin | socioeconomic status | marital status| education level | type of residency | living in rural area | health area| deprivation index | other> (C6.7)

23) Lifestyle factors included: <*select multiple:* None | tobacco use | alcohol use | frequency of exercise | other> (C6.8)

24) Percentage of the population covered by the data source in the catchment area *<number>* % (C1.11.2)

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) *<free text>* (C1.11.1)

III. Data source - Quantitative descriptors

25) Population size (total number of unique individuals with records captured in the data source): *<number>* (C7.1)

26) Population size by age: (C7.3)

<number> newborn infants (0 to 27 days),

<number> infants and toddlers (28 days to 23 months),

<number> children (2 to 11 years),

<number> adolescents (12 to 17 years),

<number> adults (18 to 45 years),

<number> adults (46 to 64 years),

<number> adults (65 to 75 years),

<number> adults (76 to 85 years),

<number> adults (86 years and over)

27) ***Active*** population size (total number of unique individuals alive and currently registered, i.e. where a record was created and not closed): *<number>* (C7.1.1)

28) ***Active*** population size by age: *(C7.3.1)*

<number> newborn infants (0 to 27 days),

<number> infants and toddlers (28 days to 23 months),

<number> children (2 to 11 years),

<number> adolescents (12 to 17 years),

<number> adults (18 to 45 years),

<number> adults (46 to 64 years),

<number> adults (65 to 75 years),

<number> adults (76 to 85 years),

<number> adults (86 years and over)

29) Median time between first and last available records for unique individuals captured in the data source: *<number>* years (B6.3)

30) Median time between first and last available records for unique ***active*** individuals (alive and currently registered) captured in the data source: *<number>* years (B6.3.1)

IV. Data source – Data flows and management

31) Governance details - Documents or webpages that describe the overall governance of the data source and processes and procedures for data capture and management, data access, data quality check and validation results, utilisation for research purposes *<free text| weblink>* (C2.3)

32) If further follow-up needed, possibility of: (tick as applicable):

☐ Accessing biospecimens (C2.13)

Biospecimen access conditions, if applicable *<free text>* (C2.13.1)

☐ Contacting patients / practitioners (C2.7)

33) The process of collection and recording of data in the data source (this could include the tools used, such as surveys) or a description of the system that the originator uses to gather data and store it the data source) *<free text>* (C4.3)

34) The trigger for creating *a record* (e.g.: hospital discharge, specialist encounter, medicinal product dispensing etc.) *<free text>* (C5.2)

35) Event triggering registration of *a person* in the data source: <*select one:* Birth | immigration| residency obtained | start of insurance coverage | disease diagnosis | start of treatment | practice registration | other> (C1.6)

If 'other', the triggering event for a person to be registered *<free text>* (C1.6.1)

36) Event triggering de-registration of a *person* in the data source <*select one*: Death | emigration | insurance coverage end | practice deregistration| loss to follow up | end of treatment | other > (C1.7)

If 'other', the triggering event for a person to be de-registered *<free text>* (C1.7.1)

37) Where the data source is created by the linkage of other data sources:

Linkage strategy<*select one:* Deterministic | probabilistic | combination | other> (B5.2)

Linkage variable *<free text>* (B5.2.1)

Linkage completeness *<free text>* (B5.3)

38) Names of linked data sources *<free text>* (B4.1)

39) The following data management specifications apply for the data source (tick as applicable):

☐ Allows data validation (e.g.: access to original medical charts) (C2.7)

☐ Records are preserved indefinitely (C8.5)

If not, records are preserved for *<number>* years (C8.5.1)

☐ Approval is needed for publishing results of a study using its data (C2.9)

40) Informed consent for use of data for research <*select one:* Not Required | required for general use | required for all studies | required for intervention studies | waiver | other> (C2.5)

If 'other', further details on the informed consent *<free text>* (C2.5.1)

41) Data source refreshed on fixed dates *<month>* (C8.2)

42) Data source last refresh *<date>* (C8.3)

43) CDM mapping: Is the data source ETL-ed to a CDM? <*select one:* Yes | No> (D1.2.1.1)

44) CDM name <*select one:* OMOP | ConcepTION | Nordic | Sentinel | PCORnet | VSD | i2b2 | CDISC SDTM | PEDSnet | other> (D1.2)

If 'other', CDM used *<free text>* (D1.2.1)

45) CDM website reference *<free text| weblink>* (D1.4)

46) CDM release frequency *<number>* months (D1.7)

47) Data source ETL status <*select one:* Planned | Completed | In progress | Not ETL-ed> (B7.1)

ETL frequency *<number>* months (B7.5)

Version(s) of CDM(s) to which the data source has been ETL-d *<free text>* (B7.3)

48) 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) *<free text| weblink>* (B7.4)

V. Data source - Vocabularies

49) Medicinal product information available: <*select multiple*: Not captured | Brand name | Batch number | Formulation | Strength | Package size| Dose | Dosage regime | Route of administration> (C6.15)

If applicable, medicinal product vocabulary used: <*select multiple:* ART 57 | IFA GmbH| EQDM | SPN | MTHSPL | Not coded (Free text) | Other> (C6.15.1)

50) Cause of death vocabulary:<*select multiple:* Not captured | ICPC | ICD9| ICD10| ICD1| Read| SNOMED | SNOMED CT | MedDRA | OPCS| CCS | EDC| Not coded (Free text)| Other> (C6.11.1)

If 'other' cause of death vocabulary used *<free text>* (C6.11.2)

51) Quality of life measurements: < *select multiple:* Not captured | AQoL-8D | QOLS | MQOL | MQOL-E | HRQOL | WHOQOL | EQ5D | 15D | SF-36 | SF-6D | HUI | Not coded (Free text)| other> (C6.28)

If 'other' Quality of life measurements or patient reported outcomes *<free text>* (C6.28.1)

52) Prescription vocabulary: < *select multiple*: Not captured | ATC | RxNorm | EphMRA | ALT | DrugBank | Not coded (Free text) | Other> (C6.13.1)

53) Dispensing vocabulary < *select multiple:* Not captured | ATC | RxNorm | EphMRA | ALT | DrugBank | Not coded (Free text)| Other> (C6.14.1)

54) Indication vocabulary < *select multiple:* Not captured | ICPC | ICD9 | ICD10 | ICD1 | Read | SNOMED | SNOMED CT | MedDRA | OPCS | CCS | EDC | Not coded (Free text) | Other> (C6.18.1)

If 'other', indication vocabulary used <free text> (C6.18.2)

55) Procedures vocabulary < *select multiple:* Not captured | ICPC | ICD9 | ICD10 | ICD1 | Read | SNOMED | SNOMED CT | MedDRA | OPCS | CCS | EDC | Not coded (Free text) | Other> (C6.22)

56) Genetic data vocabulary < *select multiple:* Not captured | OGG | FG | GO | EGO | SOPHARM | PHARE | Other> (C6.25.1)

If 'other', genetic data vocabulary used *<free text>* (C6.25.2)

57) Biomarker data vocabulary < *select multiple*: Not captured | BMO | SMASH | FOBI | Other> (C6.26.1)

58) Diagnosis/ medical event vocabulary < *select multiple:* Not captured | ICPC | ICD9 | ICD10 | ICD1 | Read | SNOMED | SNOMED CT | MedDRA | OPCS | CCS | EDC | Not coded (Free text) | Other> (C6.9.1)

2.2. Study metadata

I. Study – Administrative details

1) *(EU PAS)* Study title and acronym: *<free text>* *(F1.2)*
2) *(EU PAS)* EU PAS register number: *<number>* (F2.2)
3) *(EU PAS)* Brief description of the study *<free text>* (F10)
4) *(EU PAS)* Study status <*select one:* planned | ongoing | finalised> (F11)
5) Institution conducting the study: *<free text|lookup of institutions>* *(F1.3)*

    Study institution contact name: *<free text>* *(F1.4)*
    Study institution contact email: *<free text>* *(F1.5)*
    Additional institutions: *<free text>* (F1.7)
    (EU PAS) Primary lead investigator name *<free text>* (F12)
    (EU PAS) Primary lead investigator contact email *<free text>* (F13)
    Primary lead investigator ORCID: *<number>* *(F1.6)*

6) *(EU PAS)* Study timelines: initial administrative steps, progress reports and final report

	Planned	Actual
Date when funding contract was signed	*<date>*(F19)	*<date>* (F19.1)
Start date of data collection	*<date>* (F20)	*<date>* (F20.1)
End date of data collection	*<date>* (F2.10)	*<date>* (F2.10.1)
Start date of data analysis	*<date>* (F21)	*<date>* (F21.1)
Date of interim report, if expected	*<date>* (F22)	*<date>* (F22.1)
Date of final study report	*<date>* (F23)	*<date>* (F23.1)

7) Network conducting the study (if applicable): *<free text|lookup of networks>* (F1.8)
8) *(EU PAS)* Country where the study is conducted *<select multiple:* ISO 3166-1 country codes>
9) Source of funding: *<select multiple*: EU institutional research programme | non-EU institutional research programme | EMA | national competent authority (NCAs) | other public funding (e.g.: hospital, university) | non for-profit organisation (e.g., charity) | pharmaceutical company and other private sector | no external funding| other> (F8.8)

    If 'other', further details on the scope of the study: *<free text>* (F8.8.1)

10) Protocol link: A link to the latest version of the protocol, if published *<weblink>* (F2.1)

    Protocol document *<uploaded document>* (F2.3)

11) *(EU PAS)* Study required by a regulator *<select one:* Yes | No | I don't know> (F14)
12) *(EU PAS)* Is the study required by a Risk Management Plan (RMP) *<select one*: Not applicable | EU RMP category 1 (imposed as condition of marketing authorisation | EU RMP category 2 (specific obligation of marketing authorisation | EU RMP category 3 (required) | Non-EU RMP only> (F14.1)
13) *(EU PAS)* Regulatory procedure number (RMP Category 1 and 2 studies only): *<free text>* (F14.2)
14) *(EU PAS)* Other study registration identification numbers and URLs as applicable: *<free text>* *(F1.1.1)*

II. Study – methodological aspects

15) The study is concerning: <*select multiple*: human medicinal product | veterinary medicinal product | herbal medicinal product | medical device | disease/health condition | medical procedure | other> (F8.1)

If 'other', further details on the study topic: <*free text*> (F8.1.1)

16) Study type: <*select one:* clinical trial | non-interventional study | not applicable> (F8.2)

If 'Not applicable', further details on the study type: <*free text*> (F8.2.1)

If Study type = clinical trial

17) Clinical trial regulatory scope: <*select one:* pre-authorisation clinical trial | post-authorisation interventional clinical trial | post-authorization low-interventional clinical trial | clinical trial not subject to marketing authorization> (F8.2.2)

18) Phase of the clinical trial (where applicable): <*select one:* phase 1 | phase 2 | phase 3 | phase 4> (F8.4)

19) Clinical trial study design:

- Clinical trial randomisation: <*select one*: randomised clinical trial |non-randomised clinical trial> (F8.3)
- Clinical trial types : <*select multiple*: low-interventional clinical trial | single-arm trial | large simple trial | pragmatic clinical trial | cluster randomised trial > (F8.3.1)

If Study type = non-interventional study

20) Non-interventional study design: <*select one*: cohort| case-control| case-only| cross-sectional| ecological| cluster design| systematic review and meta-analysis| other> (F8.3.2)

If 'other' design of non-interventional study, further details: <*free text*> (F8.3.2.1)

21) Data collection methods: <*select one:* primary data collection| secondary data collection| combined primary and secondary data collection| no individual level data collected for the purpose of the study> (F8.5)

22) Scope of the study: <*select multiple*: effectiveness study (incl. comparative)| safety study (incl. comparative) | assessment of risk minimisation measure implementation or effectiveness | drug utilisation | healthcare resource utilization | disease epidemiology | patient reported outcomes | feasibility analysis | validation of study variables (exposure, outcome, covariate) | hypothesis generation (including signal detection) | method development or testing | scoping review (including literature review) | other > (F8.6)

If 'other', further details on the scope of the study: <*free text*> (F8.6.1)

23) *(EU PAS)* Study drug information - ATC <*ATC code*> (F15)

24) *(EU PAS)* Study drug information - INN <*INN list*> (F16)

25) *(EU PAS)* Study drug information - brand name <*free text*> (F17)

26) *(EU PAS)* Medical condition to be studied <*MedDRA list*> (F18)

27) *(EU PAS)* Additional medical condition(s) <*free text*> (F18.1)

28) *(EU PAS)* Population studied: A short description of the study population <*free text*> (F2.5)

29) *(EU PAS)* Population studied: Age groups: <*select multiple:* newborn infants (0 to 27 days) | infants and toddlers (28 days to 23 months) | children (2 to 11 years) | adolescents (12 to 17

years) | adults (18 to 45 years) | adults (46 to 64 years) | adults (65 to 75 years) | adults (76 to 85 years) | adults (86 years and over) | all ages> (F2.5.1)

30) *(EU PAS)* Special population of interest: *<select multiple*: immunocompromised | renal impaired | hepatic impaired | women of child-bearing age| pregnant women| lactating women| other > (F8.9)

    If 'other', further details on the population of interest: *<free text>* (F8.9.1)

31) *(EU PAS)* Population: estimated number of subjects *<number>* (F2.5.2)

32) Setting: A short description of the study setting *<free text>* (F2.11)

33) Main study objective: A short description of the study objective *<free text>* (F8.10)

34) Interventions: A short description of the study interventions *<free text>* (F2.6)

35) Comparators: A short description of the study comparators *<free text>* (F2.7)

36) Outcomes: A short description of the study outcomes *<free text>* (F2.8)

37) Study design: A brief summary of the study design *<free text>* (F2.13)

38) Data analysis plan: A brief summary of the analysis method (e.g. risk estimation, measures of risk, internal/external validity) *<free text>* (F2.12)

39) Summary of results: A brief summary of the results of the study completion (from abstract) *<free text>* (F6.3)

    Results tables *<uploaded document>* (F6.4)

40) Study publications: Peer-reviewed papers reporting the study *<AMA citation format>* (F7.1)

41) Study report: *< uploaded document|weblink>* (F7.2)

42) Study, other information: A list of URLs to other relevant resources describing the study *<weblink>* (F7.3)

VI. Study – Data management

43) *(EU PAS)* ENCePP seal *<Y/N> (F9)*

44) *(EU PAS)* ENCePP seal relevant documents:

    ☐ Conflicts of interest of investigators *<uploaded document> (F9.1)*

    ☐ Composition of steering group and observers *<uploaded document> (F9.2)*

    ☐ Signed code of conduct *<uploaded document> (F9.3)*

    ☐ Signed code of conduct checklist *<uploaded document> (F9.3.1)*

    ☐ Signed checklist for study protocols *<uploaded document> (F9.4)*

45) Number of data sources: *<number>* (F3.2)

46) Names of data sources: *<free text|lookup of data sources>* (F3.4)

47) Sources of data: <*select multiple*: clinical trial | non-interventional study | electronic healthcare records (EHR) | administrative data (e.g. claims) | drug utilisation data | drug dispensing/prescription data | disease registry | drug registry | population registry | pregnancy registry | spontaneous reporting system | laboratory data | -omics | social media | patient surveys | data from digital health wearables | expanded access program / compassionate use | published literature | other> (F8.7)

    If 'other', further details on the sources of data: *<free text>* (F8.7.1)

48) CDM mapping: Were data sources in the study ETL-ed to a CDM? *<y/n>* (F4.1)

CDM name: *<free text>* (F4.2)

CDM mapping version: *<free text>* (F4.3)

49) The following data quality specifications apply for the study (tick as applicable):

- ☐ Data characterisation conducted (F5.1)

  If 'yes', Data characterisation moment: At what stages of the study were data characterisation steps or quality checks implemented? *< select multiple*: *after data extraction*| *after extract-transform-load to a common data model*| *after creation of study variables*> (F5.2)

  If 'yes', Data characterisation details: Provide a summary description of the data characterisation or quality check process *<free text*| *weblink>* (F5.7)

  If 'yes', Data characterisation results: Provide results of the data characterisation or quality checks, such as Sentinel Common Data Model level 1-4 checks. *<uploaded document*| *weblink>* (F5.8)
- ☐ Check conformance: Was a check of the conformance of data (i.e., data are in the correct format/syntax) completed? (F5.3)
- ☐ Check completeness (F5.4)
- ☐ Check stability: Was a check of the stability of data (e.g., codes) over time completed? (F5.5)
- ☐ Check logical consistency (F5.6)

50) Procedure of data extraction: Upload or provide a link (e.g., GitHub) to the procedures, such as codes or scripts, used to extract data from the data source instance used in the study *<uploaded document*| *weblink>* (F6.1)

51) Procedure of result generation: Upload or provide a link (e.g., Github) to the procedures, such as study scripts, used to generate the study results *<uploaded document*| *weblink>* (F6.2)

2.3. Institution metadata

1) Institution name *<free text>* (A1.2)
2) Type of institution *<select one:* Academic institution | civic authority | government agency | healthcare payer | network of primary care practices | non-profit organisation | pharmaceutical industry | private organisation | public health authority | research centre | statistical authority | CRO | Other> (A1.4)
3) Institution country *<select multiple:* ISO 3166-1 country codes> (A1.5)
4) Institution website *<free text>* (A1.6)
5) Institution contact name <*free text>* (A1.7.1)
6) Institution contact email *<free text>* (A1.7)
7) Institution role *<select multiple:* Data custodian | data provider | researcher | other > (A1.8)
8) Network(s) of which the institution is a member *<free text*|*lookup of networks>* (A1.10)

2.4. Network metadata

1) Network name and acronym *<free text>* (E1.2)
2) Network country *<select multiple:* ISO 3166-1 country codes> (E1.9)
3) Network website *<free text>* (E1.4)
4) Network contact name *<free text>* (E1.5)
5) Network email contact *<free text>* (E1.6)
6) Institutions that are part of the network *<free text|lookup of institutions>* (E1.7)

Text and format conventions used: