

***** This document is for information only *****

You will need to apply online in order to add a Data Source

'Add a Data Source' Questionnaire

RWD Catalogues

Please complete the questionnaire to register your data source in the RWD Catalogues. Mandatory fields are marked with an asterisk (*).

The questionnaire comprises **16 questions** divided in **4 steps**: Administrative Details, Data Elements Collected, Quantitative Descriptors and Data Flows and Management. You agreed with the terms and conditions when you joined the RWD Catalogues.

Step 1: Administrative details

1. Administrative details

Name of data source* [Free text] *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.*

Data source acronym [Free text]

Data holder* [Drop-down menu with Institutions registered in RWD Catalogues] *Name of the institution that sustains the data source. If an institution is not included in the catalogue, it must be inserted before creating a data source record by following this link: [Add Institution in the catalogue](#). Note that a few days may be needed for the approval of the new institution in the catalogue. Once the Institution is approved, you will be able to link it to a data source and other content of the catalogue.*

Data source type* [Drop-down menu]

- Administrative healthcare claims
- Biobank
- Birth registry
- Cancer registry
- Congenital anomaly registry
- Death registry
- Diagnostic tests or procedures reimbursement
- Disease registry
- Drug registry
- Emergency care discharge records
- Exemptions from co-payment
- Hospital discharge records
- Hospital inpatient records
- Hospital outpatient visit records
- Induced terminations registry
- Other
- Pharmacy dispensing records
- Population registry
- Primary care medical records
- Registration with healthcare system
- Specialist ambulatory care records
- Spontaneous reporting of adverse drug reactions
- Vaccination registry

If 'Other' was selected in the last question, **Data source type, other** [Free text] If 'other', specify the type of data source

Main financial support [Drop-down menu] *Specify the main financial support sources for the data source in the last 3 years. Select all that apply.*

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

Care setting [Drop-down menu]

- Hospital inpatient care
- Hospital outpatient care
- Primary care – GP, community pharmacist level
- Primary care – specialist level (e.g., paediatricians)
- Secondary care – specialist level (ambulatory)
- Other

Data source qualification: *If the data source has successfully undergone a formal qualification process (e.g., from the EMA, ISO or other certifications), this should be described.* [Toggle switch]

- Yes
- No

If 'Yes' was selected in the last question, **Description of the qualification** [Free text]
Describe the qualification(s) granted.

2. Data source website

Where such an information is available, add a link to the dedicated webpage describing the data source

URL [Free text] *The format used should be: <http://example.com>*

Link text [Free text] *This is the text that will be displayed for the above URL.*

3. Contact details

Name and e-mail address provided here will be made public. A functional (organisation) contact e-mail may be provided

Main contact*

First name* [Free text]

Last name* [Free text]

Email* [Free text]

Alternate contact

If 'Add Contact' was selected:

First name* [Free text]

Last name* [Free text]

Email* [Free text]

4. Data source regions and languages

Data source countries* [Drop-down menu, includes European Union option] *Countries where the data originates. Select all that apply.*

Data source languages* [Drop-down menu] *Languages in which the records in the data source are recorded. Select all that apply.*

Data source regions [Drop-down menu] *Geographical regions that the data source covers. Select all that apply.*

5. Data source establishment

Data source established [Date with format dd/mm/yyyy] *The date when the data source was first set-up. This date can be different from the 'first collection date'.*

	First Collection	Last collection
Data source time span	[Date with format dd/mm/yyyy] ____/____/____	[Date with format dd/mm/yyyy] ____/____/____
	<i>The date when data started to be collected or extracted</i>	<i>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.</i>

6. Publications

A list of peer-reviewed papers or documents describing the data source (validation, data elements, representativity) or its use for pharmacoepidemiologic research

Data source publications

DOI or URL [Free text] *The DOI of the publication should be provided. The format used for the DOI should be <https://doi.org/10.1109/5.771073>. If a DOI is not available, the format used for a URL should be <http://example.com>*

DOI or URL display text [Free text] *The official title of the publication*

DOI or URL [Free text] *The DOI of the publication should be provided. The format used for the DOI should be <https://doi.org/10.1109/5.771073>. If a DOI is not available, the format used for a URL should be <http://example.com>*

DOI or URL display text [Free text] *The official title of the publication*

(...)

Step 2: Data elements collected

7. The data source contains the following information

Disease data: *Does the data source collect information with a focus on a specific disease? This might be a patient registry or other similar initiatives.* [Toggle switch]

- Yes
- No

If 'Yes' was selected in the last question, **Disease information collected** [Field with MedDRA codes]

If 'Yes' was selected in the last question, **Disease information collected, other** [Free text]
Disease or diseases for which information is collected, if not available in the above look-up

Rare diseases: *Are rare diseases captured? In the European Union a rare disease is one that affects no more than 5 people in 10,000.* [\[Toggle switch\]](#)

- Yes
- No

Pregnancy and/or neonates: *Does the data source collect information on pregnant women and/or neonatal subpopulation (under 28 days of age)?* [\[Toggle switch\]](#)

- Yes
- No

Hospital admission and/or discharge [\[Toggle switch\]](#)

- Yes
- No

ICU admission: *Is information on intensive care unit admission available?* [\[Toggle switch\]](#)

- Yes
- No

Cause of death [\[Toggle switch\]](#)

- Captured
- Not Captured

If 'Captured' was selected in the last question, **Cause of death vocabulary** [\[Drop-down menu\]](#)

- CCS
- CPT
- dm+d
- EDC
- HCPCS
- Human Phenotype Ontology (HPO)
- ICD
- ICD-10
- ICD-10-CM
- ICD-9
- ICD-9-CM
- ICD1

- ICPC
- ICPC-1
- ICPC-2
- MedDRA
- Not coded (Free text)
- OPCS
- Orphacode
- Orphanet Rare Disease Ontology (ORDO)
- Other
- Read
- SNOMED
- SNOMED CT

If 'Other' was selected in the last question, Cause of death vocabulary, other [Free text]
Cause of death vocabulary, if not available in the above look-up

Prescriptions of medicines [Toggle switch]

- Captured
- Not Captured

If 'Captured' was selected in the last question, Prescription vocabulary [Drop-down menu]

- ALT
- ATC
- DrugBank
- EphMRA
- not coded
- other
- RxNorm

If 'Other' was selected in the last question, Prescriptions vocabulary, other [Free text]
Prescriptions vocabulary, if not available in the above look-up

Dispensing of medicines [Toggle switch]

- Captured
- Not Captured

If 'Captured' was selected in the last question, **Dispensing of medicines vocabulary** [Drop-down menu]

- ALT
- ATC
- DrugBank
- EphMRA
- not coded
- other
- RxNorm

If 'Other' was selected in the last question, **If 'other,' what vocabulary is used?** [Free text]
Dispensing vocabulary, if not available in the above look-up

Advance therapy medicinal products (ATMP): *Is information on advanced therapy medicinal products included? A medicinal product for human use that is either a gene therapy medicinal product, a somatic cell therapy product or a tissue engineered products as defined in Regulation (EC) No 1394/2007 [Reg (EC) No 1394/2007 Art 1(1)].* [Toggle switch]

- Yes
- No

Contraception: *Is information on the use of any type of contraception (oral, injectable, devices etc.) available?* [Toggle switch]

- Yes
- No

Indication for use: *Does the data source capture information on the therapeutic indication for the use of medicinal products?* [Toggle switch]

- Captured
- Not Captured

If 'Captured' was selected in the last question, **Indication for use vocabulary** [Drop-down menu]

- CCS

- CPT
- dm+d
- EDC
- HCPCS
- Human Phenotype Ontology (HPO)
- ICD
- ICD-10
- ICD-10-CM
- ICD-9
- ICD-9-CM
- ICD1
- ICPC
- ICPC-1
- ICPC-2
- MedDRA
- Not coded (Free text)
- OPCS
- Orphacode
- Orphanet Rare Disease Ontology (ORDO)
- Other
- Read
- SNOMED
- SNOMED CT

If 'Other' was selected in the last question, **Indication vocabulary, other** [Free text]
Indication vocabulary, if not available in the above look-up

Medical devices: *Is information on medicinal devices (e.g., pens, syringes, inhalers) available?* [Toggle switch]

- Yes
- No

Administration of vaccines [\[Toggle switch\]](#)

- Yes
- No

Procedures: *Does the data source capture information on procedures (e.g., diagnostic tests, therapeutic, surgical interventions)?* [\[Toggle switch\]](#)

- Captured
- Not Captured

If 'Captured' was selected in the last question, **Procedures vocabulary** [\[Drop-down menu\]](#)

- CCS
- CPT
- dm+d
- EDC
- HCPCS
- Human Phenotype Ontology (HPO)
- ICD
- ICD-10
- ICD-10-CM
- ICD-9
- ICD-9-CM
- ICD1
- ICPC
- ICPC-1
- ICPC-2
- MedDRA
- Not coded (Free text)
- OPCS
- Orphacode
- Orphanet Rare Disease Ontology (ORDO)
- Other
- Read
- SNOMED

- SNOMED CT

If 'Other' was selected in the last question, **If 'other,' what vocabulary is used?** [Free text]
Procedures vocabulary, if not available in the above look-up

Clinical measurements: *Is information on clinical measurements (e.g., BMI, blood pressure, height) available?* [Toggle switch]

- Yes
- No

Healthcare provider: *Is information on the person providing healthcare (e.g., physician, pharmacist, specialist) available? The healthcare provider refers to individual health professionals or a health facility organisation licensed to provide health care diagnosis and treatment services including medication, surgery and medical devices.* [Toggle switch]

- Yes
- No

Genetic data: *Are data related to genotyping, genome sequencing available?* [Toggle switch]

- Captured
- Not Captured

If 'Captured' was selected in the last question, **Genetic data vocabulary** [Drop-down menu]

- EGO
- FG
- GO
- HGNC
- OGG
- Other
- PHARE
- SOPHARM

If 'Other' was selected in the last question, **Genetic data vocabulary, other** [Free text]
Genetic data vocabulary, if not available in the above look-up

Biomarker data: *Does the data source capture biomarker information? The term 'biomarker' refers to a broad subcategory of medical signs (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. [Toggle switch]*

- Captured
- Not Captured

If 'Captured' was selected in the last question, **Biomarker data vocabulary** [Drop-down menu]

- BMO
- FOBI
- HPO
- Other
- SMASH

If 'Other' was selected in the last question, **Biomarker vocabulary, other** [Free text]
Biomarker data vocabulary, if not available in the above look-up

Patient-reported outcomes: *Is information on patient-reported outcomes (e.g. quality of life) available? [Toggle switch]*

- Yes
- No

Patient-generated data: *Is patient-generated information (e.g., from wearable devices) available? [Toggle switch]*

- Yes
- No

Units of healthcare utilisation: *Are units of healthcare utilisation (e.g., number of visits to GP per year, number of hospital days) available or can they be derived? Units of healthcare utilisation refer to the quantification of the use of services for the purpose of preventing or curing health problems. [Toggle switch]*

- Yes
- No

Unique identifier for persons: *Are patients uniquely identified in the data source?* [\[Toggle switch\]](#)

- Yes
- No

Diagnostic codes [\[Toggle switch\]](#)

- Captured
- Not Captured

If 'Captured' was selected in the last question, **Diagnosis / medical event vocabulary, other** [\[Drop-down menu\]](#)

- CCS
- CPT
- dm+d
- EDC
- HCPCS
- Human Phenotype Ontology (HPO)
- ICD
- ICD-10
- ICD-10-CM
- ICD-9
- ICD-9-CM
- ICD1
- ICPC
- ICPC-1
- ICPC-2
- MedDRA
- Not coded (Free text)
- OPCS
- Orphacode
- Orphanet Rare Disease Ontology (ORDO)
- Other
- Read

- SNOMED
- SNOMED CT

If 'Other' was selected in the last question, **Diagnosis / medical event vocabulary, other** [\[Free text\]](#) *Diagnosis / medical event vocabulary, if not available in the above look-up*

Medicinal product information [\[Toggle switch\]](#)

- Captured
- Not Captured

If 'Captured' was selected in the last question, **Medicinal product information** [\[Drop-down menu\]](#)

- Active ingredient(s)
- Batch number
- Brand name
- Dosage regime
- Dose
- Formulation
- Package size
- Route of administration
- Strength

If 'Other' was selected in the last question, **Medicinal product information, other** [\[Free text\]](#) *Medicinal product information, if not available in the above look-up*

Medicinal product vocabulary [\[Drop-down menu\]](#)

- AIC
- ART 57
- ATC
- CNF
- DIN
- dm+d

- EQDM
- Gemscript
- GTIN
- IFA GmbH
- MTHSPL
- NDC
- NDF
- Not coded (Free text)
- Other
- RxNorm
- SNOMED
- SPN
- WHO Drug
- Z-index (G-standard)

If 'Other' was selected in the last question, Medicinal product vocabulary, other [\[Free text\]](#) *Medicinal product vocabulary, if not available in the above look-up*

Quality of life measurements [\[Toggle switch\]](#)

- Captured
- Not Captured

If 'Captured' was selected in the last question, Quality of life measurements vocabulary [\[Drop-down menu\]](#)

- 15D
- AQoL-8D
- EQ5D
- HRQOL
- HUI
- MQOL
- MQOL-E
- Not Captured

- Not coded (Free text)
- Other
- QOLS
- SF-36
- SF-6D
- WHOQOL

If 'Other' was selected in the last question, **Quality of life measurements, other** [Free text]
Quality of life measurements vocabulary, if not available in the above look-up

Lifestyle factors [Toggle switch]

- Captured
- Not Captured

If 'Captured' was selected in the last question, **Lifestyle factors included** [Drop-down menu]

- Alcohol use
- Diet
- Frequency of exercise
- Not capture
- Other
- Tobacco use

If 'Other' was selected in the last question, **Lifestyle factors included other** [Free text]
Lifestyle factors included, if not available in the above look-up

Sociodemographic information [Toggle switch]

- Captured
- Not Captured

If 'Captured' was selected in the last question, **Sociodemographic information** [Drop-down menu]

- Age
- Country of origin

- Deprivation index
- Education level
- Ethnicity
- Gender
- Health area
- Living in rural area
- Marital status
- Not Captured
- Other
- Pharmaceutical copayment
- Sex
- Socioeconomic status
- Type of residency

If 'Other' was selected in the last question, **Sociodemographic information other** [[Free text](#)] *Sociodemographic information, if not available in the above look-up*

Step 3: Quantitative Descriptors

8. Population Qualitative Data

Population age groups [[Drop-down menu](#)] *Population age groups captured in the data source*

- Adolescents
- Adult animal
- Adults
- Adults (18 to < 46 years)
- Adults (46 to < 65 years)
- Adults (65 to < 75 years)
- Adults (75 to < 85 years)
- Adults (85 years and over)
- Adults and elderly

- All
- Children
- Egg
- Elderly
- Embryonated egg
- In utero
- Infants and toddlers
- Neonate
- Paediatric Population
- Prepubertal children
- Preterm newborn infants
- Pubertal and postpubertal adolescents
- Term newborn infants
- Young animal
- Young animal/Newborn

Estimated percentage of the population covered by the data source in the catchment area [Free text, limited by 2000 characters] *This field can also be used to include details and/or clarifications about the population and the population size that cannot be captured in other fields.*

Description of the population covered by the data source in the catchment area whose data are not collected (e.g., people who are registered only for private care) [Free text, limited by 2000 characters] *This field can also be used to include details and/or clarifications about the population and the population size that cannot be captured in other fields.*

9. Family linkage

Family linkage available in the data source permanently or can be created on an ad-hoc basis [Drop-down menu]

- None
- Permanently
- Adhoc

If 'Permanently' was selected in the last question, Family linkage available between the following persons [Drop-down menu]

- Father-child
- Household
- Mother-child
- Sibling

10. Population

Population size [Numerical value] *Total number of unique individuals with records captured in the data source (most recent count)*

Active population size [Numerical value] *Total number of unique individuals alive and currently registered with active records in the data source, i.e. a record was created and not closed (most recent count). For example, an active population for administrative healthcare data refers to the collection of patients for which there is an active record in the practice, i.e. the record was created and not closed because the patient moved or died.*

Population by age group *Total number of unique individuals with records captured in the data source (most recent count), split by age groups. If the population size by age cannot be described (e.g., exact numbers are not available), the field 'Estimated percentage of the field population covered by the data source in the catchment area' can be used to add further information.*

If 'Add Age group dataset' button was selected:

Age group* [Drop-down menu]

- Adolescents
- Adult animal
- Adults

- Adults (18 to < 46 years)
- Adults (46 to < 65 years)
- Adults (65 to < 75 years)
- Adults (75 to < 85 years)
- Adults (85 years and over)
- Adults and elderly
- All
- Children
- Egg
- Elderly
- Embryonated egg
- In utero
- Infants and toddlers
- Neonate
- Paediatric Population
- Prepubertal children
- Preterm newborn infants
- Pubertal and postpubertal adolescents
- Term newborn infants
- Young animal
- Young animal/Newborn

Population size [Numerical value]

Active population size [Numerical value]

11. Median observation time

Median observation time (years) [Numerical value] *The median time, in years, between first and last available records for unique individuals captured in the data source*

Median observation time active (years) [Numerical value] *The median time, in years, between first and last available records for unique active individuals (alive and currently registered) captured in the data source*

Step 4: Data flows and management

12. Access and validation

Please be aware that the uploaded document will be made public.

Documents or links to webpages that describe the overall governance, processes and procedures for data capture and management, data access, data quality check and validation results, utilisation for research purposes.

Governance details (file)

If 'Add Media' button was selected:

Add Files [Select a file (or several files) from your computer] *Unlimited number of files can be uploaded to this field. 20 MB limit. Allowed types: pdf.*

Governance details (link)

URL [Free text] *The format used for a URL should be http://example.com*

URL display text [Free text] *This is the text that will be displayed for the above URL*

Biospecimen access: *Are biospecimens available in the data source (e.g., tissue samples)?* [Toggle switch]

- Yes
- No

If 'Yes' was selected in the last question, **Biospecimen access conditions** [Free text, limited by 2000 characters]

Access to subject details: *Can individual patients/practitioners/practices included in the data source be contacted?* [\[Toggle switch\]](#)

- Yes
- No

Description of data collection [\[Free text, limited by 2000 characters\]](#) *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 in the data source.*

13. Event triggering registration

Event triggering creation of a record in the data source [\[Free text, limited by 500 characters\]](#) *The event triggering the creation of a record in the data source should be described (e.g., hospital discharge, specialist encounter, dispensation of a medicinal product, recording of a congenital anomaly). This refers in general to the creation of a record in the data source and not to the registration of a person.*

Event triggering de-registration of a person in the data source [\[Drop-down menu\]](#)

- Death
- Emigration
- End of treatment
- Insurance coverage end
- Loss to follow up
- Other
- Practice deregistration

If 'Other' was selected in the last question, **Event triggering de-registration of a person in the data source, other** [\[Free text\]](#)

Event triggering registration of a person in the data source [Drop-down menu]

- Birth
- Disease diagnosis
- Immigration
- Insurance coverage start
- Other
- Practice registration
- Residency obtained
- Start of treatment

If 'Other' was selected in the last question, **Event triggering registration of a person in the data source, other** [Free text]

14. Data source linkage

Linkage: *Is the data source described created by the linkage of other data sources (prelinked data source) and/or can the data source be linked to other data source on an ad-hoc basis?* [Toggle switch]

- Yes
- No

Linkage description, pre-linked [Free text] *Description of the linkages that are currently available in the data (i.e. created by linkage of other data sources.)*

Linkage description, possible linkage [Free text] *Linkages to other data sources that are possible on an ad-hoc basis.*

Linked data sources

If 'Add Linked data source' button is selected:

Pre-Linked *Is the data source described created by the linkage of other data sources?*

Data source [Drop-down menu with Data Sources registered in RWD Catalogues]

Data source, other [Free text]

Linkage strategy [Drop-down menu]

- None
- Combination
- Deterministic
- Probabilistic
- Other

Linkage variable [Free text]

Linkage completeness [Free text]

15. Data management specifications that apply for the source

Possibility of data validation: *Can validity of the data in the data source be verified (e.g., access to original medical charts)?* [Toggle switch]

- Yes
- No

Data source preservation: *Are records preserved in the data source indefinitely?* [Toggle switch]

- Yes
- No

If 'No' was selected in the last question, **Data source preservation length (years)** [Free text] *Where the records are not indefinitely preserved, the number of years for which the records are kept*

Approval for publication: *Is an approval needed for publishing the results of a study using the data source?* [Toggle switch]

- Yes
- No

Informed consent for use of data for research [Drop-down menu] *Is informed consent required to use the data for research purposes?*

- Not Required
- Required for general use
- Required for all studies
- Required for intervention studies
- Waiver
- Other

If 'Other' was selected in the last question, **Informed consent, other** [Free text] *The type of informed consent required when it does not fit the above categories*

Data source refresh [Drop-down menu] *The data source refresh months, if the data source is refreshed on fixed months, or time intervals between two consecutive refreshes (select all that apply).*

- January
- February
- March
- April
- May
- June
- July
- August

- September
- October
- November
- December
- Quarterly
- Every 6 months
- Yearly

Data source last refresh [Date with format dd/mm/yyyy] *The date when the last refresh of the data source occurred*

16. Common Data Model (CDM) mapping

CDM mapping: *Has the data source been converted (ETL-ed) to a common data model?* [Toggle switch]

- Yes
- No

CDM Mappings

If 'Add Data source CDM Mapping' button was selected:

CDM name [Drop-down menu] *Name by which the CDM is most frequently referred*

- ConcepTION
- Other CDM
- OMOP
- Sentinel Common Data Model
- CDISC SDTM
- PCORnet CDM

CDM name (other) [Free text] *The CDM name, if not available in the above look-up*

Data source ETL CDM version [Free text] *The version of CDM to which the data source has been converted*

Data source ETL status [Drop-down menu] *The status of the conversion process (ETL) of the data source to the CDM*

- Planned
- Completed
- In Progress
- Not ETL-ed

Data source ETL frequency [Free text] *The frequency of the conversion process (ETL) in months*

Data source ETL specifications (file) [Select a file from your computer] *Documents describing the mapping of the data source to the CDM, including codes and scripts to convert original data to the CDM. One file only. 20 MB limit. Allowed types: pdf.*

Data source ETL specifications (link) [Free text] *URLs describing the mapping of the data source to the CDM, including codes and scripts to convert original data to the CDM. This must be an external URL such as http://example.com.*

***** End of questionnaire *****